

PCT
 WELTORGANISATION FÜR GEISTIGES EIGENTUM
 Internationales Büro
 INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
 INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)



| | | |
|--|-----------|--|
| (51) Internationale Patentklassifikation ⁶ : C12N 15/54, 9/12, 15/11, 15/85, C12Q 1/68, A01K 67/027 | A2 | (11) Internationale Veröffentlichungsnummer: WO 99/33998 (43) Internationales Veröffentlichungsdatum: 8. Juli 1999 (08.07.99) |
|--|-----------|--|

(21) Internationales Aktenzeichen: **PCT/EP98/08216**

(22) Internationales Anmeldedatum: **22. Dezember 1998 (22.12.98)**

(30) Prioritätsdaten:
 197 57 984.1 24. Dezember 1997 (24.12.97) DE

(71) Anmelder (für alle Bestimmungsstaaten ausser US): **BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE).**

(72) Erfinder; und
 (75) Erfinder/Anmelder (nur für US): **HAGEN, Gustav [DE/DE]; Bertha-von-Suttner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Roggen-dorfstrasse 59, D-51061 Köln (DE).**

(74) Gemeinsamer Vertreter: **BAYER AKTIENGE-SELLSCHAFT; D-51368 Leverkusen (DE).**

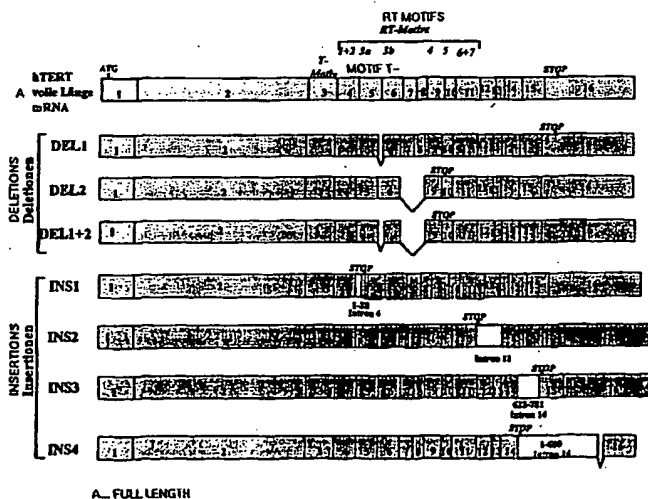
(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht
Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.

(54) Title: **REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB-UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF**

(54) Bezeichnung: **REGULATORISCHE DNA-SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOMERASE-UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG**

BEST AVAILABLE COPY



(57) Abstract

The present invention relates to regulatory DNA sequences containing promotor sequences, in addition to intervening sequences, for the human catalytic telomerase sub-unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.

(57) Zusammenfassung

Diese Erfindung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

| | | | | | | | |
|----|--------------|----|------------------------|----|-----------|----|-----------|
| AL | Albanien | ES | Spanien | LS | Lesotho | SI | Slowenien |
| AM | Armenien | FI | Finnland | LT | Litauen | SK | Slowakei |
| AT | Österreich | FR | Frankreich | LU | Luxemburg | SN | Senegal |
| AU | Australien | GA | Gabun | LV | Lettland | SZ | Swasiland |
| AZ | Aserbaidshan | GB | Vereinigtes Königreich | MC | Monaco | TD | Tschad |

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen
Telomerase-Untereinheit und deren diagnostische und therapeutische Verwen-
dung

5 Aufbau und Funktion der Chromosomenenden

Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten
10 Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).

15 Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.

20 Desweiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomer enthielten (Sandell und Zakian, 1993).

25 Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der

30 Okazaki Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-

werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

5

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 10
1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

15

Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

Das Enzym Telomerase synthetisiert die Telomere

20

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzellern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994).

25

Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet

26 ... ist im Zellkern lokalisiert. Während der RNA-

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

10 Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomeraseaktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten CD34⁺38⁺-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumor-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgenese ist.

5 Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomersequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase
10 enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein. Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomersequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus von Tumorzellen zur Stabilisation ihrer Telomere.
15

Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzwelligen Strahlen schädigen nicht nur die Tumorzellen,
20 sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgeweben wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle
25 Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden
30 dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapie von Krebs dar.

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

5

Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und post-translationalale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

15

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein“) erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Sp1 („specificity protein1“) spezifisch gebunden werden.

20

25

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter“), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements“) be-

30

schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie vermitteln gewebes- oder zellspezifische Promotoreigenschaften.

Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

15

Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

20

Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen

Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon
5 beinhaltet.

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere
10 weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

15 Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).
20

Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

25 Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN- α , IFN- β , IFN- γ , TNF, TNF- α , TNF- β , Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls
30 zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1 α , - β), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF, M-CSF.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispielsweise:

Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bakterielle Nitroreductase, bakterielle β -Glukuronidase, pflanzliche β -Glukuronidase aus *Secale cereale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle Carboxypeptidase, bakterielle β -Lactamase, bakterielle Cytosindeaminidase, humane Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phosphatase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-Sequenzen zählen auch andere Blutgerinnungsfaktoren

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen beispielsweise:

Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC), β -Galaktosidase (β -Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes Wachstumshormon (hGH), β -Glukuronidase (GUS), Grün-fluoreszierendes Protein (GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.

Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch
5 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch
10 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen,
15 zen, sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen.

Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziiertes Virus, Herpes Simplex Virus, Vaccina Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.
20

Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfindungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem.
25

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

Die Zubereitungen der erfindungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-
30

turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

5 Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilsequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.

10 Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF β , Insulin oder Insulin-like Growth Factor (ILGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.

15 Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intravenös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, transdermal, intraperitoneal, als Aerosol oder oral) am Ort eines Tumors oder die systemische Applikation dieser Zubereitungen.

20

Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.

25 Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.

Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

30

A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,

B. Messung des Substanzeffektes auf die Reportergenexpression.

10 Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.

15 Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.

20 Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der 5'-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilfragmenten unterschiedlichster Länge als Sonde.

25 Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.

Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

- 5 A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reportergen mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten;
- 10 C. Vergleich des diagnostischen Werts mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;

15 Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assoziierten Zustand, der wiederum einen pathogenen Zustand indiziert.

Erläuterung der Abbildungen:

20 Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies

25 A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Sprague Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).

30 B: Zu Fig.1 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

Fig. 2: Restriktionsanalyse der rekombinanten λ -DNA des Phagenklons P12, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

5 Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene λ -DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die
10 Spuren 3 - 7 enthalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des rekombinanten Phagens. Die Pfeile kennzeichnen die zwei λ -Arme des Vektors EMBL3 Sp6/T7.

Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten λ -DNA des Phagenklons, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.
15

20 A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene λ -DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur
25 14: nicht verdaut.

30 B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5'-hTC-cDNA Fragment eingesetzt.

Fig. 4: Partielle DNA-Sequenz der 5'-flankierenden Region und des Promotors vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

5

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

10

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 5'GTTAAGTTGTAGCTTACACTGGTTCTC 3' benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

15

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die dargestellte Sequenz entspricht SEQ ID NO 2.

20

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5'- und 3'-flankierende Region.

25

Exons sind als durchnummerierte schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymschnittstellen sind wie folgt gekennzeichnet: S, SacI; X, XhoI. Die relative Anordnung der fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch

30

die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

5

10

15

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

20

25

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgespligte 5'- und 3'-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

30

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.

Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3'-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5' Primer in der kodierenden Region oder der 5'-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promotors.

5 11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenzabschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.

10

Fig. 11: Aktivität des hTC-Promotors in HEK-293 Zellen.

Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-flankierenden hTC Genregion schematisch dargestellt. Das ATG-Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der hTC Promotor-Luziferase Konstrukte dargestellt. Das promotorlose pGL2-Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbildungsseite ist die relative Luziferaseaktivität der verschiedenen Promotorkonstrukte in HEK-Zellen als durchgehende Balken gezeigt. Die Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten durchgeführt wurden.

15

20

25 Tab. 1: Exon Intron Übergänge des hTC-Gens

30

Aufgelistet sind die Nukleotidsequenzen an den 3'- und 5' Spliceübergängen des hTC-Gens. Die Konsensussequenzen für Donor und Akzeptorsequenzen (AG und GT) sind durch graue Rechtecke unterlegt. Intronsequenzen (Kleinbuchstaben) und Exonsequenzen (Großbuchstaben), die die Spliceakzeptor- und Donorstellen flankieren sind gezeigt. Die Größe der Exons und Introns ist in bp angegeben.

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

5 Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in

10 Intron 2.

Tab. 1

| 3' Acceptor Sequence | | | | | 5' Donor Sequence | | | | |
|----------------------|-----------------------|----------|------|----|----------------------|------|----------------------|---------|-------|
| Intron | Exon | Exon No. | bp | bp | Intron | Exon | Intron | Intr on | bp |
| | | | | | | | | No. | No. |
| 5' flanking Region | | | | | | | | | |
| caggcgcttcccccgag | GTTCAGGCAGCGCTGCGT | 1 | 281 | | CGCCCCCTCCTTCCGCCAG | | gtggcctccccgggtcg | 1 | 104 |
| catgtccttctcgtttaag | GTGTCTGCTGCTGAAGAGC | 2 | 1354 | | TGGCTGCCGAGGAGCCCCAG | | gtgaggaggtggtggccgt | 2 | 8616 |
| gaggggtctctctattgag | GGTTGGCTGTGTTCCGGC | 3 | 196 | | TGCAAGCATTTGAATCAG | | gtactgtatccccacgcca | 3 | 2089 |
| cccatgtgtcccccgag | ACAGCACTTGAAGAGGTG | 4 | 181 | | GTCCGCAGAGAAAAAGAGG | | gtggctgtgctttgttta | 4 | 687 |
| ctgcctccactcacagag | GCCGAGCGTCTCACCTCGA | 5 | 180 | | TGAGCTGTACTTTGTCAAG | | gtgggtgccccgggaccccc | 5 | 494 |
| ccctctcctctgccccgag | GTGGATGTGACGGGCGCGT | 6 | 156 | | CAAGGCCTTCAAGAGCCAC | | gtaaggttcacgtgtgata | 6 | >4660 |
| ctccccgtctgttttcgag | GTCTCTACCTTGACAGACC | 7 | 96 | | TGCCGTGTCATCGAGCAG | | gtctgggcaactgccccgca | 7 | 980 |
| ctgtgtcttccccgccccag | AGTCCTCCTCCCTGAATGAG | 8 | 86 | | CCGTGCGCATCAGGGGCAA | | gtgagtcaggtggccaggt | 8 | 2484 |
| gtattttcccttatttttag | GTCTCTGCTGCTGTTGGTG | 9 | 114 | | CGGGGATTGCGCGGGACGG | | gtgaggcctcctctctcccc | 9 | 1984 |
| cattgccccctctgccttag | GACCTGCTGCTGCTGTTGGTG | 10 | 72 | | ACGCGAAAACTTCCTCAG | | gtgaggccccgtgccccgtg | 10 | 1871 |
| attccccccctgtgtctdag | CTATGCCCGGACCTCCATC | 11 | 189 | | TGCAGAGCGACTACTCCAG | | gtgagcgacacctggccgga | 11 | 3804 |
| tctttcttggcgactctag | GTGAACAGCCTCCAGACGG | 12 | 127 | | CCTGTTTCTGGATTGTCAG | | gtgagcaggctgatggtca | 12 | 880 |
| ctgtccgccatctctctdag | GTTCACCGCATGTGTCTG | 13 | 62 | | TCCTGCTGCAGCGGTACAG | | gtgagccccaccacaagggg | 13 | 3187 |
| agcctctgttttcccccdag | GGATGTGCTGCGTGGGGCCAA | 14 | 125 | | CTGAAAGCCCAAGAACGCAG | | gtatgtgcagggtgcctggc | 14 | 781 |
| tctgatttttggcccccgag | CCCAGACGCAGCTGAGTCG | 15 | 138 | | CTGGGCTCACTCAGGACAG | | gcaagtgtgggtggaggcc | 15 | 536 |
| | | 16 | 664 | | TTTTTCAGTTTGTGAAAAAA | | 3' flanking Region | | |

Tab. 2

| Faktoren | Lokalisation in Intron 2 |
|-------------------------|--|
| C/EBP | 2925 |
| CRE.2 | 2749 |
| Sp1 | 2378, 4094, 4526, 4787, 4835, 4995 |
| AP-2 CS3 | 5099 |
| AP-2 CS4 | 2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496 |
| AP-2 CS5 | 5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498 |
| PEA3 | 934, 2505 |
| P53 | 2125 |
| GR uteroglobin | 848, 1487, 2956 |
| PR uteroglobin | 3331 |
| Zeste-white | 1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055 |
| GRE | 846 |
| MyoD-MCK right site/rev | 447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557 |
| MyoD-MCK left site | 108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030 |
| Ets-1 CS | 6408 |
| API | 3784, 4406 |
| CREB | 2801 |
| GATA-1 | 839, 1390, 3154 |
| c-Myc | 108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030 |
| CACCC site | 991 |
| CCAAT site | 1224 |
| CCAC box | 992 |
| CAAT site | 463, 2395 |
| Rb site | 992, 4663 |
| TATA | 3650 |
| CDEI | 106, 1564, 1606, 1690, 1732, 1816, 1900, 1984 |

Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5' und 3' liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen bzw. eventuell auch ghTC-Pseudogene existieren.

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschrte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

5 Beispiel 2

Zur Isolierung der 5' flankierenden hTC-Gensequenz wurden ca $1,5 \times 10^6$ Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 µm; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa
10 500 bp langen 5'-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 µg/ml
15 Heringsperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybridisierung über Nacht wurde die Prähybridisierungslösung mit $1,5 \times 10^6$ cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte
20 durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 als potentiell positiv heraus. Eine λ-DNA Präparation dieses Phagens Ausubel *et al.*
25 (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen
30 Experimenten jeweils 1 bis $1,5 \times 10^6$ Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmustert.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3' hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach λ -DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umklontiert (Beispiel 4).

Beispiel 3

Um zu untersuchen, ob auch das 5'-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde λ -DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5'-Bereich hybridisiert (Fig. 3).

Da die isolierte λ -DNA des positiven Klons auch mit dem extremen 5'-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5'-Sequenzbereich.

Beispiel 4

Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einem das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.

Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umkloniert. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5'-flankierenden des ghTC-Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5' Sequenz.

Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen XhoI und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes XhoI-BamHI-Fragment sowie ein 1,8 kb großes BamHI-Fragment subkloniert. Der Kombinations-Restriktionsverdau mit den Enzymen XhoI und XbaI führte zur Klonierung von einem 6,5 kb großen XhoI-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem XhoI-Fragment.

Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym XhoI. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes XhoI-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subkloniert.

Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI subkloniert. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes XhoI-Subfragment erhalten.

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenklons P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie XhoI-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

5

Zur Klonierung des 3' von Phagenklon P17 und 5' von Phagenklon P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers AP1 (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt.

10 Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested“ PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested“ genspezifischen Primers und 10 pmol des „nested“ Marathon Adaptor Primers AP2 (5'-

15 ACTATAGGGCACGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen

20 kloniert.

25

30

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTCTTGAAGGCCTTG-3') sowie der „nested“ genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATAACC-3') zusammen mit der HDL-Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten
5 Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested“ genspezifischen Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4 ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking
10 Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-GGTGGATGTGACGGGCGCGTACG-3') und des „nested“ genspezifischen Primers C5K-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses
Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen genomischen hTC-Region identifiziert (s. Fig. 7).

15

Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software (DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten
20 Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5 zusammengesetzt. Überlappende Phagenklonbereiche sind in Fig. 7 schematisch
25 dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2 unterstrichen.

Contig1:

| | | | | | | | | |
|----|------------|-------------|------------|------------|------------|-------------|------------|------|
| | ACTTGAGCCC | AAGAGTTCAA | GGCTACGGTG | AGCCATGATT | GCAACACCAC | ACGCCAGCCT | TGGTGACAGA | 70 |
| 5 | ATGAGACCCT | GTCTCAAAAA | AAAAAAAAAA | AATTGAAATA | ATATAAAGCA | TCTTCTCTGG | CCACAGTGGA | 140 |
| | ACAAAACCAG | AAATCAACAA | CAAGAGGAAT | TTTGAAAAC | ATACAAACAC | ATGAAAAATTA | AACAATATAC | 210 |
| | TTCTGAATGA | CCAGTGAGTC | AATGAAGAAA | TAAAAAGGA | AATTGAAAAA | TTTATTTAAG | CAATGATAA | 280 |
| | CGGAAACATA | ACCTCTCAAA | ACCCACGGTA | TACAGCAAAA | GCAGTGCTAA | GAAGGAAGTT | TATAGCTATA | 350 |
| | AGCAGCTACA | TCAAAAAAGT | AGAAAAGCCA | GGCGCAGTGG | CTCATGCCTG | TAATCCCAGC | ACTTTGGGAG | 420 |
| 10 | GCCAAGGCGG | GCAGATCGCC | TGAGGTCAGG | AGTTCGAGAC | CAGCCTGACC | AACACAGAGA | AACCTTGTCG | 490 |
| | CTACTAAAAA | TACAAAATTA | GCTGGGCATG | GTGGGCATG | CCTGTAATCC | CAGCTACTCG | GGAGGCTGAG | 560 |
| | GCAGGATAAC | CGCTTGAACC | CAGGAGGTGG | AGGTTGCGGT | GAGCCGGGAT | TGCGCCATTG | GACTCCAGCC | 630 |
| | TGGGTAACAA | GAGTGAAACC | CTGTCTCAAG | AAAAAAAAAA | AAGTAGAAAA | ACTTAAAAAT | ACAACCTAAT | 700 |
| | GATGCACCTT | AAAGAACTAG | AAAAGCAAGA | GCAAACTAAA | CCTAAAATTG | GTAAGAAGAA | AGAAATAATA | 770 |
| | AAGATCAGAG | CAGAAATAAA | TGAACTGAA | AGATAACAAT | ACAAAAGATC | AACAAAATTA | AAAGTTGGTT | 840 |
| 15 | TTTTGAAAAG | ATAAACAAAA | TTGACAAACC | TTTGCCCAGA | CTAAGAAAAA | AGGAAAGAAG | ACCTAAATAA | 910 |
| | ATAAAGTCAG | AGATGAAAAA | AGAGACATTA | CAACTGATAC | CACAGAAATT | CAAGGATCA | CTAGAGGCTA | 980 |
| | CTATGAGCAA | CTGTACACTA | ATAAATTGAA | AAACCTAGAA | AAAATAGATA | AATTCTTAGA | TGCATACAAC | 1050 |
| | CTACCAAGAT | TGAACCATGA | AGAAATCCAA | AGCCCAACAA | GACCAATAAC | AATAATGGGA | TTAAAGCCAT | 1120 |
| 20 | AATAAAAAGT | CTCCTAGCAA | AGAGAAGCCC | AGGACCCAAT | GGCTTCCCTG | CTGGATTTTA | CCAATCATTT | 1190 |
| | AAAGAAGAAT | GAATTCCAAT | CCTACTCAAA | CTATTCTGAA | AAATAGAGGA | AAGAATACTT | CCAACTCAT | 1260 |
| | TCATCATGGC | CAGTATTACC | CTGATTCCAA | AACCAACAAA | AAACACATCA | AAAACAAACA | ACAAACAAAA | 1330 |
| | CAGAAAGAAA | GAAACTACA | GGCCAATATC | CCTGATGAAT | ACTGATACAA | AAATCCTCAA | CAAAACACTA | 1400 |
| | GCAAACCAAA | TTAAACAACA | CCTTCGAAAG | ATCATTCAAT | GTGATCAAGT | GGGATTATT | CCAGGGATGG | 1470 |
| 25 | AAGGATGGTT | CAACATATGC | AAATCAATCA | ATGTGATACA | TCATCCCAAC | AAATGAAGT | ACAAAACTA | 1540 |
| | TATGATTATT | TCACTTTATG | CAGAAAAAGC | ATTTGATAAA | ATTCTGCACC | CTTCATGATA | AAAACCTCA | 1610 |
| | AAAAACCAGG | TATACAAGAA | ACATACAGGC | CAGGCACAGT | GGCTCACACC | TGCGATCCCA | GCACCTGGG | 1680 |
| | AGGCCAAGGT | GGGATGATTG | CTTGGGCCCC | GGAGTTTGAG | ACTAGCCTGG | GCAACAAAA | GAGACCTGGT | 1750 |
| | CTACAAAAAA | CTTTTTTAAA | AAATTAGCCA | GGCATGATGG | CATATGCCTG | TAGTCCAGC | TAGCTGGAG | 1820 |
| 30 | GCTGAGGTGG | GAGAATCACT | TAAGCCTAGG | AGGTCGAGGC | TGCAGTGAGC | CATGAACATG | TCAGTGACT | 1890 |
| | CCAGCCTAGA | CAACAGAACA | AGACCCCACT | GAATAAGAAG | AAGGAGAAGG | AGAAGGGAGA | AGGGAGGGAG | 1960 |
| | AAGGATGGAG | GAGGAGAAGG | AGGAGGTGGA | GGAGAAGTGG | AAGGGGAAGG | GGAAGGGAAA | GAGGAAGAAG | 2030 |
| | AAGAAACATA | TTTCAACATA | ATAAAGCCCC | TATATGACAG | ACCGAGGTAG | TATTATGAGG | AAAAACTGAA | 2100 |
| | AGCCTTTCTT | CTAAGATCTG | GAAAAAGACA | AGGGCCCACT | TTCAACCATG | TGATTCAACA | TAGTACTAGA | 2170 |
| 35 | AGTCCATAGT | AGAGCAATCA | GATAAGAGAA | GAATAATAAA | GGCATCCCAA | CTGGAAGGA | AGAAGTCAAA | 2240 |
| | TTATCCTGTT | TGCAGATGAT | ATGATCTTAT | ATCTGGAAAA | GACTTAAGAC | ACCACTAAAA | AACATTATGA | 2310 |
| | GCTGAAATTT | GGTACAGCAG | GATACAAAA | CAATGTACAA | AAATCAGTAG | TATTTCTATA | TTCCAACAGC | 2380 |
| | AAACAATCTG | AAAAAGAAAC | CAAAAAAGCA | GCTAGATAAA | AAATTAACA | GCTAGGAATT | AACCAAGAAA | 2450 |
| | GTGAAAGATC | TCTACAATGA | AACTATAAAA | ATGTTGTATA | AAGAAATTGA | AGAGGGCACA | AAAAAAGAAA | 2520 |
| 40 | AGATATTCCA | TGTTTCATAGA | TTGGAAGAAT | AAATCTGTTT | AAATGTCCA | TACTACCCAA | AGCAATTTAC | 2590 |
| | AAATTCAATG | CAATCCCTAT | TAAATACTA | ATAGCGTTCT | TCACAGAAAT | AGAAGAAACA | ATTCTAAGAT | 2660 |
| | TTGTACAGAA | CCACAAAAGA | CCCAGAATAG | CCAAAGCTAT | CCTGACCAAA | AGAAGACAAA | CTGGAAGCAT | 2730 |
| | CACATTACCT | GACTTCAAA | TATACTACAA | AGCTATAGTA | ACCCAAACTA | CATGGTACTG | GCATAAAAC | 2800 |
| | AGATGAGACA | TGGACCAGAG | GAACAGAATA | GAGAATCCAG | AAACAAATCC | ATGCATCTAC | AGTGAACCTA | 2870 |
| 45 | TTTTTGACAA | AGGTGCCAAG | AACATACTTT | GGGGAAGAGA | TATCTCTTTC | AATAAATGGT | AGTGAGGAAA | 2940 |
| | CTGGATATCC | ATATGCAAAA | TAACAATACT | AGAACTCTGT | CTCTCACCAT | ATACAAAAGC | AAATCAAAAT | 3010 |
| | GGATGAAAGG | CTTAAATCTA | AAACCTCAAA | CTTTGCAACT | ACTAAAAGAA | AACACCGGAG | AAACTCTCCA | 3080 |
| | GGACATGGGA | TGTGGCAAA | ACTTCTTGAG | TAATTCCTGT | CAGGCACAGG | CAACCAAGG | CAAAACAGC | 3150 |
| | AAATGGGATC | ATATCAAGTT | AAAAAGCTTC | TGCCCAGCAA | AGGAAACAAT | CAACAAAGAG | AAGAGACAAC | 3220 |
| 50 | CCACAGAATG | GGAGAATATA | TTTGCAAACT | ATTACTCTAA | CAAGGAATTA | ATAACAGTA | TATATAAGGA | 3290 |
| | GCTCAAACTA | CTCTATAAGA | AAAACACATA | ATAAGCTGAT | TTTCAAAAA | AAGCAAAAGA | TCTGGGTAGA | 3360 |
| | CATTTCTCAA | AATAAGTCAT | ACAAATGGCA | AACAGGCATC | TGAAAATGTG | CTCAACACCA | CTGATCATCA | 3430 |
| | GAGAAATGCA | AATCAAAACT | ACTATGAGAG | ATCATCTCAT | CCAGTTTAAA | ATGGCTTTTA | TTCAAAAGAC | 3500 |
| | AGGCAATAAC | AAATGCCAGT | GAGGATGTGG | ATAAAAGGAA | ACCCCTGGAC | ACTGTTGGTG | GGAATGGAAG | 3570 |
| 55 | TTGTATACC | TATGGAGAAC | AGTTTGAAG | TTCTCAAAA | AACTAAAAT | AAAGCTACCA | TACAGCAATC | 3640 |
| | CCATTGCTAG | GTATATACTC | CAAAAAGGG | AATCAGTGTA | TCAACAAGCT | ATCTCCACTC | CCACATTTAC | 3710 |
| | TGCAGCACTG | TTCATAGCAG | CCAAGGTTTG | GAAGCAACCT | CAGTGTCAT | CAACAGACGA | ATGAAAAAAG | 3780 |
| | AAAATGTGGT | GCACATACAC | AATGGAGTAC | TACGCAGCCA | TAAAAAAGAA | TGAGATCCTG | TCAGTTGCAA | 3850 |
| | CAGCATGGGG | GGCACTGGTC | AGTATGTTAA | GTGAAATAAG | CCAGGCACAG | AAAGACAAAC | TTTTCATGTT | 3920 |
| 60 | CTCCCTTACT | TGTGGGAGCA | AAAATTAATA | CAATTGACAT | AGAAATAGAG | GAGAATGGTG | GTTCTAGAGG | 3990 |
| | GGTGGGGGAC | AGGGTGACTA | GAGTCAACAA | TAATTTATTG | TATGTTTTAA | AATAACTAAA | AGAGTATAAT | 4060 |
| | TGGGTTGTTT | GTAACACAAA | GAAAGGATAA | ATGCTTGAAG | GTGACAGATA | CCCCATTTAC | CCTGATGTGA | 4130 |
| | TTATTACACA | TTGTATGCCT | GTATCAAAAT | ATCTCATGTA | TGCTATAGAT | ATAAACCTTA | CTATATTAAA | 4200 |
| | AATTAATAAT | TTAATGGCCA | GGCACGGTGG | CTCATGTCCG | TAATCCCAGC | ACTTTGGGAG | GCCGAGGCGG | 4270 |
| 65 | GTGGATCACC | TGAGGTCAGG | AGTTTGAAC | CAGTCTGGCC | ACCATGATGA | AACCCTGTCT | CTACTAAAGA | 4340 |
| | TACAAAAAAT | AGCCAGGCGT | GGTGGCACAT | ACCTGTAGTC | CCAACTACTC | AGGAGGCTGA | GACAGGAGAA | 4410 |
| | TTGCTTGAAC | CTGGGAGGCG | GAGGTTGCAG | TGAGCCGAGA | TCATGCCACT | GCAGTCGAGC | CTGGGTGACA | 4480 |
| | GAGCAAGACT | CCATCTCAAA | ACAAAACAAA | AAAAAAGAA | ATTAATAATG | TAATTTTTAT | GTACCGTATA | 4550 |
| | AATATATACT | CTACTATATT | AGAAGTTAAA | AATTAACA | ATTATAAAG | GTAATTAACC | ACTTAATCTA | 4620 |
| 70 | AAATAAGAAC | AATGTATGTG | GGGTTTCTAG | CTTCTGAAGA | AGTAAAAAGT | ATGGCCACGA | TGGCAGAAAT | 4690 |
| | GTGAGGAGGG | AACAGTGGAA | GTTACTGTTG | TAGACGCTC | ATACTCTCTG | TAAGTGACTT | AAATTTAACC | 4760 |
| | AAAGACAGGC | TGGGAGAAGT | TAAAGAGGCA | TTCTATAAGC | CCTAAAACAA | CTGCTAATAA | TGGTGAAAGG | 4830 |
| | TAATCTCTAT | TAATTACCAA | TAATTACAGA | TATCTCTAAA | ATCGAGCTGC | AGAATTGGCA | CGTCTGATCA | 4900 |
| | CACCGTCTCT | TCATTACCG | TGCTTTTTTT | CTTGTGTGCT | TGGAGATTTT | CGATTGTGTG | TTCTGTGTTG | 4970 |
| 75 | GTAAACTTAA | ATCTGTATGA | ATCTGAAAC | GAAAATGGT | GGTGATTTC | TCCAGAAGAA | TTAGATACC | 5040 |
| | TGGCAGGAAG | CAGGTGGCTC | TGTGGACCTG | AGCCACTTCA | ATCTTCAAGG | GTCTCTGGCC | AAGACCCAGG | 5110 |

| | | | | | | | | |
|----|------------|------------|------------|------------|------------|-------------|-------------|-------|
| | TGCAAGGCAG | AGGCCTGATG | ACCCGAGGAC | AGGAAAGCTC | GGATGGGAAG | GGGCGATGAG | AAGCCTGCCT | 5180 |
| | CGTTGGTGAG | CAGCGCATGA | AGTGCCCTTA | TTTACGCTTT | GCAAAGATTG | CTCTGGATAC | CATCTGGAAA | 5250 |
| | AGGGCGCCAG | CGGGAATGCA | AGGAGTCAGA | AGCCTCTCTG | TCAAACCCAG | GCCAGCAGCT | ATGGCGCCCA | 5320 |
| 5 | CCCGGGCGTG | TGCCAGAGGG | AGAGGAGTCA | AGGCACCTCG | AAGTATGGCT | TAAATCTTTT | TTTCACTTGA | 5390 |
| | AGCAGTGACC | AAGGTGTATT | CTGAGGGAAG | CTTGAGTTAG | GTGCCTTCTT | TAAAACAGAA | AGTCATGGAA | 5460 |
| | GCACCCCTCT | CAAGGGAAAA | CCAGACGCC | GCTCTGCGGT | CATTACCTC | TTTCTCTCT | CCCTCTCTTG | 5530 |
| | CCCTCGCGGT | TTCTGATCGG | GACAGAGTGA | CCCCGCTGGA | GCTTCTCCGA | GCCCCTGCTG | AGGACCTCT | 5600 |
| | TGCAAGGGC | TCCACAGACC | CCCGCCCTGG | AGAGAGGAGT | CTGAGCCTGG | CTTAATAACA | AACTGGGATG | 5670 |
| | TGGCTGGGG | CGGACAGCGA | CGGCGGGATT | CAAAGACTTA | ATTCCATGAG | TAAATTCAAC | CTTTCCACAT | 5740 |
| 10 | CCGAATGGAT | TTGGATTTTA | TCTTAATATT | TTCTTAAATT | TCATCAAATA | ACATTTCAGGA | CTGCAGAAAT | 5810 |
| | CCAAAGGCGT | AAAACAGGAA | CTGAGCTATG | TTTGCCAAGG | TCCAAGGACT | TAATAACCAT | GTTTCAGAGG | 5880 |
| | ATTTTTCGCC | CTAAGTACTT | TTTATTGGTT | TTTATAAGGT | GGCTTAGGGT | GCAAGGGAAA | GTACACGAGG | 5950 |
| | AGAGGCCTGG | GCGGCAGGGC | TATGAGCAGC | GCAGGGCCAC | CGGGGAGAGA | GTCCTCCGCC | TGGGAGGCTG | 6020 |
| 15 | ACAGCAGGAC | CACTGACCGT | CCTCCCTGGG | AGCTGCCACA | TTGGGCAACG | CGAAGGCGGC | CACGCTGCGT | 6090 |
| | GTGACTCAGG | ACCCCATACC | GGCTTCTCTG | GCCCACCCAC | ACTAACCAGG | GAAGTCACGG | AGCTCTGAAC | 6160 |
| | CCGTGGAAC | GAACATGACC | CTTGCTGGC | TGCTTCCCTG | GGTGGGTCAA | GGGTAATGAA | GTGGTGTGCA | 6230 |
| | GGAAATGGCC | ATGTAAATTA | CACGACTCTG | CTGATGGGGA | CCGTTCCCTC | CATCATTATT | CATCTTCACC | 6300 |
| | CCCAGGACT | GAATGATTCC | AGCAACTTCT | TCCGGTGTGA | CAAGCCATGA | CAAAACTCAG | TACAAACACC | 6370 |
| 20 | ACTCTTTTAC | TAGGCCACA | GAGCACGGSC | CACACCCCTG | ATATATTAA | AGTCCAGGAG | AGATGAGGCT | 6440 |
| | GCTTTTCAGC | ACCAGGCTGG | GGTGACAACA | CGCGCTGAAC | AGTCTGTTC | TCTAGACTAG | TAGACCCTGG | 6510 |
| | CAGGCACTCC | CCAGATTCT | AGGGCCTGGT | TGCTGCTTCC | CGAGGGCGCC | ATCTGCCCTG | GAGACTCAGC | 6580 |
| | CTGGGCTGCC | ACACTGAGGC | CAGCCCTGTC | TCCACACCTC | CCGCCCTCAG | GCCTCAGCTT | CTCCAGCAGC | 6650 |
| | TTCTTAAACC | TGGGTGGGG | CGTGTTCAG | CGCTAGTGT | TCACCTGTCC | CACCTGTGCT | TGCTCAGCTG | 6720 |
| 25 | ACGTAGCTCG | CAGGTTTCT | CCTCACATGG | GGTGTCTGTC | TCCTTCCCA | ACACTCACAT | GCGTTGAAGG | 6790 |
| | GAGGAGATT | TGCGCCTCCC | AGACTGGCTC | CTCTGAGCCT | GAACCTGGCT | CGTGGCCCCC | GATGCAGGTT | 6860 |
| | CCTGGCGTCC | GGCTGCACGC | TGACCTCCAT | TTCCAGCGCG | TCCCGTCTC | CTGTCTCTG | CCGGGGCTG | 6930 |
| | CCGGTGTGTT | CTTCTGTTT | TGTGCTCCT | TCCACGTCCA | GCTGCGTGTG | TCTCTGCCCG | CTAGGGTCTC | 7000 |
| | GGGGTTTTTA | TAGGCATAGG | ACGGGGGCGT | GGTGGGCCAG | GGCGCTCTTG | GGAAATGCAA | CATTGGGTG | 7070 |
| 30 | TGAAAGTAGG | AGTGCCTGTC | CTCACCTAGG | TCCACGGGCA | CAGGCCTGGG | GATGGAGCCC | CCGCCAGGGA | 7140 |
| | CCCGCCCTTC | TCTGCCCAGC | ACTTTCCTGC | CCCCCTCCCT | CTGGAACACA | GAGTGGCAGT | TCCACAAGC | 7210 |
| | ACTAAGCATC | CTCTTCCCAA | AAGACCCAGC | ATTGGCACCC | CTGGACATTT | GCCCCACAGC | CCTGGGAATT | 7280 |
| | CACGTGACTA | CGCACATCAT | GTACACACTC | CCGTCCACGA | CCGACCCCGG | CTGTTTTATT | TAAATAGCTA | 7350 |
| | CAAAGCAGGG | AAATCCCTGC | TAAAATGTCC | TTTAAACAA | TGGTTAAACA | AACGGGTCCA | TCCGCACGGT | 7420 |
| 35 | GGACAGTTCC | TCACAGTGAA | GAGGAACATG | CCGTTTATA | AGCCTGCAGG | CATCTCAAGG | GAATTACGCT | 7490 |
| | GAGTCAAAAC | TGCCACCTCC | ATGGGATACG | TACGCAACAT | GCTCAAAAAG | AAAGAAATTC | ACCCCATGGC | 7560 |
| | AGGGGAGTGG | TAGGGGGGT | TAAGGACGGT | GCTGGGGGCA | GCTGGGGGCT | ACTGCACGCA | CCTTTACTA | 7630 |
| | AAGCCAGTTT | CCTGGTTCTG | ATGGTATTGG | CTCAGTTATG | GGAGACTAAC | CATAGGGGAG | TGGGGATGGG | 7700 |
| | GGAACCCGGA | GGCTGTGCCA | TCTTTGCCAT | CCCCGAGTGT | CCTGGGCAGG | ATAATGCTCT | AGAGATGCC | 7770 |
| 40 | ACGTCCTGAT | TCCCCAAAC | CTGTGGACAG | AAACCGCCCG | CTTTGCAAGT | CTTGTCTCCG | GTGATCTCCG | 7840 |
| | TGAGGACCTT | GAGGTCTGGG | ATCCTTCGGG | ACTACCTGCA | GGCCCGAAAA | GTAATCCAGG | GGTCTGGGA | 7910 |
| | AGAGGCGGGC | AGGAGGGTCA | GAGGGGGGCA | CGCTCAGGAC | GATGGAGGCA | GTCAGTCTGA | GGGTGAAAG | 7980 |
| | GGAGGGAGGG | CTCGAGCCC | AGGCCGTGCA | GGCCTCCAG | AAGCTGGAAG | AAGCGGGGAA | GGGACCTCC | 8050 |
| | ACGGAGCCTG | CAGCAGGAAG | GCACGGCTGG | CCCTTAGCCC | ACCAGGGCCC | ATCGTGGACC | TCCGGCCTCC | 8120 |
| 45 | GTGCCATAGG | AGGGCACTCG | CGCTGCCCTT | CTAGCATGAA | GTGTGTGGGG | ATTTGCAGAA | GCAACAGGAA | 8190 |
| | ACCCATCCGT | TGTGAATCTA | GGATTATTTT | AAAACAAAGG | TTTACAGAAA | CATCCAAGGA | CAGGCTGAA | 8260 |
| | GTGCCCTCCG | GCAAGGGCAG | GGCAGGCACG | AGTGAATTTT | TTTAGCTATT | TATTTTATT | TACTTACTTT | 8330 |
| | CTGAGACAGA | GTTATGCTCT | TGTTGCCCGC | GCTGGAGTGC | AGCGGCATGA | TCTTGGCTCA | CTGCAACCTC | 8400 |
| | CGTCTCCTGG | GTTCAAGCAA | TTCTCGTGCC | TACGCCCTCC | AAGTAGCTGG | GATTTTCAGG | GTGCACCACC | 8470 |
| 50 | ACACCCGGCT | AATTTTGTAT | TTTATAGTAG | GATGGGCTTT | CACCATGTTG | GTCAGCTGTA | TCTCAAAATC | 8540 |
| | CTGACCTCAG | GTGATCCGCC | CACCTCAGCC | TCCCAAAGTG | CTGGGATTAC | AGGCATGAGC | CACCTGCACCT | 8610 |
| | GGCCTATTTA | ACCATTTTAA | AACTTCCCTG | GGCTCAAGTC | ACACCCACTG | GTAAGGAGTT | CATGGAGTTC | 8680 |
| | AATTTCCCTG | TACTCAGGA | GTTACCTCTC | TTTGTATTTT | TCTGTAATTC | TTCGTAGACT | GGGGATACAC | 8750 |
| | CGTCTCTTGA | CATATTCACA | GTTTCTGTGA | CCACCTGTTA | TCCCATGGGA | CCCACTGCAG | GGGCACTGG | 8820 |
| | GAGGCTGAGC | GCTTCAGGTC | CCAGTGGGGT | TGCCATCTGC | CAGTAGAAAC | CTGATGTAGA | ATCAGGGGCC | 8890 |
| 55 | AAGTGTGGAC | ACTGTCCCTG | ATCTCAATGT | CTCAGTGTGT | GCTGAAACAT | GTAGAAATTA | AAGTCCATCC | 8960 |
| | CTCCTACTCT | ACTGGGATTG | AGCCCCCTCC | CTATCCCCCC | CCAGGGGCAG | AGGAGTTTCT | CTCACTCCTG | 9030 |
| | TGGAGGAAGG | AATGATACTT | TGTTATTTTT | CACTGCTGGT | ACTGAATCCA | CTGTTTCATT | TGTTGGTTTG | 9100 |
| | TTTGTTTTGT | TTTGAGAGGC | GGTTTCACTC | TTGTTGCTCA | GGCTGGAGGG | AGTGCAATGG | CGCGATCTTG | 9170 |
| 60 | GCTTACTGCA | GCCTCTGCCT | CCCAGGTTCA | AGTGATTCTC | CTGCTTCCGC | CTCCCATTTG | GCTGGGATTA | 9240 |
| | CAGGCACCCG | CCACCATGCC | CAGCTAATTT | TTTGTATTTT | TAGTAGAGAC | GGGGGTGGGT | GGGGTTCACC | 9310 |
| | ATGTTGGCCA | GGCTGGTCTC | GAACCTCTGA | CCTCAGATGA | TCCACCTGCC | TCTGCCCTCT | AAAGTGCTGG | 9380 |
| | GATTACAGGT | GTAGGCCACC | ATGCCACAGT | CAGCAATTTT | TCTGTTTAGA | AACATCTGGG | TCTGAGGTAG | 9450 |
| | GAAGCTCACC | CCACTCAAGT | GTTGTGGTGT | TTTAAAGCAA | TGATAGAATT | TTTTTATTGT | TGTTAGAACA | 9520 |
| 65 | CTCTTGATGT | TTTACACTGT | GATGACTAAG | ACATCATCAG | CTTTTCAAAG | ACACACTAAC | TGCACCATTA | 9590 |
| | ATACTGGGGT | GTCTTCTGGG | TATCAGCAAT | CTTCAATGAA | TGCCGGGAGG | CGTTTCTCTG | CCATGCACAT | 9660 |
| | GGTGTAAATT | ACTCCAGCAT | AATCTTCTGC | TTCCATTTCT | TCTCTTCCCT | CTTTTAAAT | TGTGTTTTCT | 9730 |
| | ATGTTGGCTT | CTCTGCAGAG | AACCACTGTA | AGCTACAAC | TAACTTTTGT | TGGAACAAAT | TTTCCAAACC | 9800 |
| | GCCCTTTTGC | CTTAGTGCCA | GAGACAATTC | ACAAACACAG | CCCTTAAAA | AGGCTTAGGG | ATCACTAAGG | 9870 |
| 70 | GGATTCTAG | AAGAGCGACC | TGTAATCCTA | AGATTTTACA | AGACGAGGCT | AACCTCCAGC | GAGCTGACA | 9940 |
| | GCCAGGAGG | GGTGCGAGGC | CTGTTCAAAT | GCTAGCTCCA | TAAATAAAGC | AATTTCTCTC | GGCAGTTTCT | 10010 |
| | GAAGCTAGGA | AAGGTTACAT | TTAAGGTTGC | TTTGTGTAGC | ATTTTCAGTG | TTGCCGACCT | CAGCTACAGC | 10080 |
| | ATCCCTGCAA | GCCCTCGGGA | GACCCAGAAG | TTTCTCGCCC | CCTTAGATCC | AAACTTGAGC | AACCTCAGGT | 10150 |
| | CTGGATTCTT | GGGAAGTCTT | CAGCTGTCTT | GCGGTTGTGC | CGGGGCCCA | GGTCTGGAGG | GGACCACTGG | 10220 |
| 75 | CCGTGTGGCT | TCTACTGCTG | GGCTGGAAGT | CGGGCTCTCT | AGCTCTGCAG | TCCGAGGCTT | GAGGCCAGGT | 10290 |
| | GCCTGGACCC | CGAGGCTGCC | CTCCACCCTG | TCCGGGCGGG | ATGTGACCAG | ATGTTGGCCT | CCATGCCCAG | 10360 |
| | ACAGAGTGCC | GGGGCCAGG | GTCAAGGCCG | TTGTGGCTGG | TGTGAGGCGC | CCGGTGCCTG | GCCAGCAGGA | 10430 |
| | GCGCTGGCT | CCATTTCCCA | CCCTTTCTCG | ACGGGACCGC | CCCGGTGGGT | GATTAACAGA | TTTGGGGTGG | 10500 |

| | | | | | | | | |
|----|------------|-------------|-------------|-------------|------------|------------|------------|-------|
| | TTTGCTCATG | GTGGGGACCC | CTCGCCGCT | GAGAACCTGC | AAAGAGAAAT | GACGGGCTCG | TGTCAAGGAG | 10570 |
| | CCCAAGTCGC | GGGGAAAGTGT | TGCAGGGAGG | CACCTCCGGGA | GGTCCCGCGT | GCCTGTCAG | GGAGCAATGC | 10640 |
| | GTCTTCGGGT | TCGTCCCTCAG | CGCGTCTAC | GCCTCCCTTC | CCTCCCTTC | ACGTCCGGCA | TTCGTGGTGC | 10710 |
| 5 | CCGGAGCCCG | ACGCCCCGCG | TCCGGACCTG | GAGGCAGCCC | TGGGTCTCCG | GATCAGGCCA | GCGGCCAAAG | 10780 |
| | GGTCGCCGCA | CGCACCTGTT | CCCAGGGCCT | CCACATCATG | GCCCTCCCT | CGGGTTACCC | CACAGCCTAG | 10850 |
| | GCCGATTGCA | CCTCTCTCCG | CTGGGGCCCT | CGCTGGCGTC | CCTGCACCTT | GGGAGCGCGA | GCGGCGCGCG | 10920 |
| | GGCGGGGAAG | CGCGGCCAG | ACCCCGGGT | CCGCCGGAG | CAGCTGCCT | GTGGGGCCA | GGCCGGGCTC | 10990 |
| | CCAGTGGATT | CGCGGCCACA | GACGCCAGG | ACCGGCTCC | CCAGTGGCG | GAGGGACTGG | GGACCCGGGC | 11060 |
| 10 | ACCCGTCTG | CCCCTTACCC | TTCCAGCTCC | GCCTCTCCG | CGCGGACCCC | GCCTGTCCTC | GACCCCTCCC | 11130 |
| | GGGTCCCGCG | CCCAGCCCC | TCCGGGCCCT | CCAGGCCCT | CCCCTTCTT | TCCGCGGCC | CGCCCTCTCC | 11200 |
| | TCCGCGCGCG | AGTTTCAGGC | AGCGCTGCGT | CCTGCTGCGC | ACGTGGGAAG | CCCTGGCCCC | GGCCACCCCC | 11270 |
| | GCGATCCCG | GCGCTCCCG | CTGCCGAGCC | GTGCGCTCCC | TGCTGCGCAG | CCACTACCCG | GAGGTGCTGC | 11340 |
| | CGCTGGCCAC | GTTCTGTCCG | CGCTGGGGC | CCCAGGGCTG | GCGGTGCTG | CAGCGCGGGG | ACCCGGCGGC | 11410 |
| 15 | TTTCCGCGCG | CTGGTGGCCC | AGTGCTGGT | GTGCTGCCCC | TGGGACGCAC | GGCGGCCCCC | CGCCGCCCCC | 11480 |
| | TCCTTCCGCC | AGGTGGGCT | CCCCGGGGT | GGCGTCCCG | TGGGGTTAG | GGCGGCCGGG | GGGAACCCAG | 11550 |
| | GACATGCGGA | GAGCAGCGCA | GGCGACTCAG | GGCGTTTCC | CCGAGGTGT | CCTGCTGAA | GGAGCTGGTG | 11620 |
| | GCCCGAGTGC | TGCAGAGGCT | GTGCGAGCGC | GGCGGAGA | ACGTGCTGGC | CTTCGGCTTC | CGCGTGTGG | 11690 |
| | ACGGGGCCCG | CGGGGGCCCC | CCCAGGGCCT | TCACCAACAG | CGTGGCAGC | TACCTGCCCA | ACACGGTGAC | 11760 |
| 20 | CGACGCACTG | CGGGGGAGCG | GGGCGTGGG | GCTGCTGCTG | CGCCGCGTGG | GCGACGACGT | GCTGGTTCAC | 11830 |
| | CTGCTGGCAG | GCTGCGCGCT | CTTTGTGCTG | GTGGCTCCCA | GCTGCGCCTA | CCAGGTGTGC | GGCGCGCCGC | 11900 |
| | TGTACAGACT | CGGCGCTGCC | ACTCAGGCC | GGCCCCGCC | ACACGCTAGT | GGACCCCGAA | GGCGTCTGGG | 11970 |
| | ATGCGAACGG | GCCTGGAACC | ATAGCGTCAG | GGAGGCGGGG | GTCCCCCTGG | GCCTGCCAGC | CCCGGGTGGC | 12040 |
| | AGGAGGCGCG | GGGGCAGTGC | CAGCCGAAGT | CTGCCGTTG | CCAAGAGGCC | CAGGCGTGGC | GCTGCCCTTG | 12110 |
| 25 | AGCCGGAGCG | GACGCCCGTT | GGGAGGGGT | CCTGGGCCCA | CCCGGGCAGG | ACGCGTGGAC | CGAGTGACCG | 12180 |
| | TGGTTTCTGT | GTGGTGTAC | CTGCCAGACC | CGCCGAAGAA | GCCACCTCTT | TGGAGGGTGC | GCTCTCTGGC | 12250 |
| | ACCGCCCACT | CCACCCATC | CGTGGGCCG | CAGCAGCAGC | CAGGCCCCCC | ATCCACATCG | CGCCCAACCA | 12320 |
| | GTCCCTGGGA | CACGCCCTGT | CCCCCGGTGT | ACGCCGAGAC | CAAGCACTTC | CTCTACTCCT | CAGGCGACAA | 12390 |
| | GGAGCAGCTG | CGGCCCTCCT | TCCTACTCAG | CTCTCTGAGG | CCAGCCTGA | CTGGCGCTCG | GAGGCTCGTG | 12460 |
| 30 | GAGACCATCT | TTCTGGGTTT | CAGGCCCTGG | ATGCCAGGGA | CTCCCCGCG | GTGCCCGCC | CTGCCCGGAC | 12530 |
| | GCTACTGGCA | AATGCGGCCC | CTGTTTCTGG | AGCTGCTTGG | GAACCAACGC | CAGTGCCCC | ACGGGTGCTG | 12600 |
| | CCTCAAGACG | CACGTGCCCG | TGCCAGCTGC | GGTACCCCA | GCAGCCGGTG | TCTGTGCCCG | GGAGAAGCCC | 12670 |
| | CAGGGCTCTG | TGGCGGCCCC | CGAGGAGGAG | GACACAGACC | CCCGTCGCTT | GSTGCAGCTG | CTCCGCCAGC | 12740 |
| 35 | ACAGACCCCT | CTGGCAGGTG | TACGGCTTCG | TGCGGCCCTG | CCTGCCCGCG | CTGGTGCCCC | CAGGCTCTGT | 12810 |
| | GGGCTCCAGG | CACAACGAAC | GCCGCTTCCT | CAGGAACACC | AAGAAGTTCA | TCTCCCTGGG | GAGCATGCC | 12880 |
| | AAGCTCTCGC | TGCAGGAGCT | GACGTGGAAG | ATGAGCGTGC | GGGACTGCGC | TTGGCTGCGC | AGGAGCCAG | 12950 |
| | GTGAGGAGCT | GGTGGCCGTC | GAGGGGCCAG | CCCCAGTGC | TGAATGACGT | AGGGGCTCAG | AAAAGGGGCG | 13020 |
| | AGGCAGAGCC | CTGGTCTCTC | TGTCTCCATC | GTCACGTGGG | CACACGTGGC | TTTTCGCTCA | GGACGTGCG | 13090 |
| | TGGACACCGT | GATCTCTGCC | TCTGCTCTCC | CTCTGTCCCA | GTTTGACATA | ACTTACGAGG | TTACCTTCA | 13160 |
| 40 | CGTTTGTATG | GACACGCGGT | TTCCAGGCGC | CGAGGCCAGA | GCAGTGAACA | GAGGAGGCTG | GCGCGGCGAG | 13230 |
| | TGGAGCCGGG | TTGCCGGCAA | TGGGGAGAAG | TGTCTGGAAG | CACAGACGCT | CTGGCGAGGG | TGCCTGCAGG | 13300 |
| | TTACCTATAA | TCCTCTTCGC | AATTTCAAGG | GTGGGAATGA | GAGGTGGGGA | CGAGAACCCC | CTCTTCTCTG | 13370 |
| | GGGTGGGAGG | TAAGGGTTTT | GCAGGTGCAC | TGTGTCAGCC | AATATGCAGG | TTTGTGTTTA | AGATTAAAT | 13440 |
| 45 | GTGTGTTGAC | GGCCAGGTGC | GGTGGCTCAC | GCCGGTAATC | CCAGCACTTT | GGGAAGCTGA | GGCAGGTGGA | 13510 |
| | TCACCTGAGG | TCAGGAGTTT | GAGACCAGCC | TGACCAACAT | GGTGAAACCC | TATCTGTACT | AAAAATACAA | 13580 |
| | AAATTAGCTG | GGCATGGTGG | TGTGTGCCCT | TAATCCCAGC | TACTTGGGAG | GCTGAGGCAG | GAGAATCACT | 13650 |
| | TGAACCCAGG | AGGCGGAGGC | TGCAGTGAGC | TGCAGTTGTG | CCATTGTACT | CCAGCCTGGG | CGACAAGACT | 13720 |
| | GAAACTCTGT | CTTTAAAAAA | AAAAAGTGTT | CGTTGATTGT | GCCAGGACAG | GTAGAGGGA | GGGAGATAAG | 13790 |
| | ACTGTTCTCC | AGCACAGATC | CTGGTCCCAT | CTTTAGGTAT | GAAGAGGGCC | ACATGGGAGC | AGAGGACAGC | 13860 |
| 50 | AGATGGCTCC | ACCTGCTGAG | GAAGGGACAG | TGTTGTGGG | TGTTACGGGG | ATGGTGCTGC | TGGCGCTGCT | 13930 |
| | CGTGTCCTCA | CCCTGTTTTT | CTGGATTGTA | TGTTGAGGAA | CCTCCGCTCC | AGCCCCCTTT | TGGCTCCAG | 14000 |
| | TGCTCCCAAG | CCCTACCGTG | GCAGTAGAA | GAAGTCCCGA | TTTACCCCC | TCCCCACAAA | CTCCCAAGAC | 14070 |
| | ATGTAAGACT | TCCGCGCATG | CAGACAAGGA | GGGTGACCTT | CTTGGGGCTC | TTTTTTTCTT | TTTTTTCTTT | 14140 |
| | TTATGGTGGC | AAAAGTCATA | TAACATGAGA | TTGGCACTCC | TAACACCGTT | TTCTGTGTAC | AGTGCAGAAT | 14210 |
| 55 | TGCTAACTCG | GCGGTGTTTA | CAGCAGGTTG | CTTGAATATG | TGCGTCTTGC | GTGACTGGAA | GTCCCTACCC | 14280 |
| | ATCGAACGGC | AGCTGCCTCA | CACCTGCTGC | GGCTCAGGTG | GACCACGCGC | AGTCAGATAA | GCGTCATGCA | 14350 |
| | ACCCAGTTTT | GCTTTTTTGT | CTCCAGCTTC | CTTCTGTGAG | GAGAGTTTGA | GTTCTCTGAT | CAGGACTCTG | 14420 |
| | CCTGTCTATT | CTGTCTCTG | ACTTCAGATG | AGGTCACAAT | CTGCCCTGG | CTTATGACAG | GAGTGAGGCG | 14490 |
| | TGGTCCCCCG | GTGTCCCTGT | CACGTGCAGG | GTGAGTGAGG | CGTTGCCCCC | AGGTGTCCCT | GTCACGTGTA | 14560 |
| 60 | GGGTGAGTGA | GGGCGGGCCC | CCGGGTGTCC | CTGTCCCGTG | CAGCGTGATT | GAGGTGTGGC | CCCCGGGTGT | 14630 |
| | CCCTGTACAG | TGTAGGGTGA | GTGAGGCGCC | ATCCCCGGGT | GTCCCTGTCA | CGTGTAGGGT | GAGTGAGGCG | 14700 |
| | TGGTCCCCCG | GTGTCCCTGT | CCCGTGCAGG | GTGAGTGAGG | CACGTGTCCC | GGGTGTCCCT | GTCACGTGCA | 14770 |
| | GGGTGAGTGA | GCGCGGGTCC | CCGGGTGTCC | CTCTCAGGTG | TAGGGTGAGT | GAGGCGCGGC | CCCAGGGTGT | 14840 |
| | CCCTGTACAG | TGTAGGGTGA | GTGAGGCGCC | GTCCCTGGGT | GTCCCTCCCA | GGTATAGGGT | GAGTGAGGCA | 14910 |
| 65 | CTGTCCCCCG | GTGTCCCTGT | CACGTGCAGG | GTGAGTGAGG | CGCGGCCCCC | GGGTGTCCCT | CTCAGGTGCA | 14980 |
| | GGGTGAGTGA | GGCGCTGTCC | CTGGGTGTCC | CTGTCTGCTG | TAGGGTGAGT | GAGGCTCTGT | CCCCAGGTGT | 15050 |
| | CCTTGGCGTT | TGCTCACTTG | AGCTTGCTCC | TGAATGTTTG | CTCTTTCTAT | AGCCACAGCT | GCGCGGGTTG | 15120 |
| | CCCATTCGCT | GGGTAGATGG | TGCAGGCGCA | GTGCTGGTCC | CCAAGCCTAT | CTTTCTGAT | GCTCGGCTCT | 15190 |
| | TCTTGGTCA | CTCTCCGTTT | CATTTTGCTA | CGGGGACACG | GGACTGCAGG | CTCTCGGCTC | CGCGTGCCCA | 15260 |
| 70 | GGCAGTCAG | CCACAGCTTC | AGGTCCGCTT | GCCTCTGCTG | GGCCTGGCTT | GCTCACCACG | GCCCGGCCAC | 15330 |
| | ATGCATGCTG | CAAATACTCC | TCTCCAGCT | TGTCTCATGC | CGAGGCTGGA | CTCTGGGCTG | CCTGTGCTGT | 15400 |
| | CTGCCACGTC | TTGCTGGAGA | CATCCACAGAA | AGGGTTCTCT | GTGCCCTGAA | GGAAAGCAAG | GCAACCCAGC | 15470 |
| | CCCTCACTT | GTCTGTGTTT | CTCCCAAGCT | TTGGCCCTCT | TGGGTGGGTG | TGGGTGGGTG | TCACACGCTT | 15540 |
| | TCACCTTATT | CTGGGCACCT | GCCGCTCATT | GCTTAGGCTG | GGCTCTGCCT | CCAGTCGCCC | CCTCACATGC | 15610 |
| | ATTGACGTC | AGCCACAGGT | TGGAGTGTCT | CTGTCTGTCT | CCTGCTCTGA | GACCCACGTC | GAGGGCCGGT | 15680 |
| 75 | GTCTCCGCGA | GCCTTCGTCA | GACTTCCCTC | TTGGGTCTTA | GTTTTGAATT | TCACTGATT | ACCTCTGACG | 15750 |
| | TTTCTATCTC | TCCATTGTAT | GCTTTTCTT | TTTATTCTT | TTTATTCTT | TCTAGCTTC | TTAGTTTATG | 15820 |
| | CATGCCTTTC | CCTCTAAGTG | CTGCCCTTACC | TGCACCTGT | GTTTTGATGT | GAAGTAATCT | CAACATCAGC | 15890 |

| | | | | | | | | |
|----|------------|-------------|------------|-------------|-------------|-------------|-------------|-------|
| | CACTTTCAAG | TGTTCTTAAA | ATACTTCAAA | GTGTTAATAC | TTCTTTTAAAG | TATTCTTATT | CTGTGATTTT | 15960 |
| | TTTCTTTGTG | CACGCTGTGT | TTTGACGTGA | AATCATTTTG | ATATCAGTGA | CTTTTAAAGTA | TTCTTTAGCT | 16030 |
| | TATTCTGTGA | TTTCTTTGAG | CAGTGAAGTA | TTTGAACACT | GTTTATGTTC | AAGATATGTA | GAGTATCAAG | 16100 |
| 5 | ATACGTAGAG | TATTTTAAAGT | TATCATTTTA | TTATTGATTT | CTAACTCAGT | TGTGTAGTGG | TCTGTATAAT | 16170 |
| | ACCAATTATT | TGAAGTTTGC | GGAGCCTTGC | TTTGTGATCT | AGTGTGTGCA | TGGTTTCCAG | AACTGTCCAT | 16240 |
| | TGTAAATTGT | ACATCCTGTC | AATAGTGGGC | ATGCATGTTT | ACTATATCCA | GCTTATTAAG | GTCCAGTGCA | 16310 |
| | AAGCTTCTGT | CTCCTTCTAG | ATGCATGAAA | TTCCAAGAAC | GAGGCCATAG | TCCCTCACCT | GGGGGATGGG | 16380 |
| | TCTGTTCAAT | TCTTCTCGTT | TGGTAGCATT | TATGTGAGGC | ATTGTTAGGT | GCATGCACGT | GGTAGAATTT | 16450 |
| | TTATCTTCTT | GATGAGTGAA | TCITTTGGAG | ACTTCTATGT | CTCTAGTAAT | CTAGTAATTC | TTTTTTTAAA | 16520 |
| 10 | TTGCTCTTAG | TACTGCCACA | CTGGGCTTCT | TTTGATTAGT | ATTTTCCTGC | TGTGTCTGTT | TTCTGCCTTT | 16590 |
| | AATTTATATA | TATATATATA | TTTTTTTTTT | TTTTGAGACA | GAGTCTTGGT | CTGTGCGCCA | GGGTGAGTGC | 16660 |
| | AGTGGTGTGA | TCACAGGTCA | GTGTAACITT | TACCTTCTGG | CCTGAGCCGT | CCTCTCACCT | CAGCCTCCTG | 16730 |
| | AGTAGCTGGA | ACTGCAGACA | CGCACCGCTA | CACCTGGCTA | ATTTTTAAAT | TTTTTCTGGA | GACAGGGTCT | 16800 |
| | TGCTGTGTTG | CCCAGGCTGG | TCTCAAATC | TTGGACTCAA | GGGATCCATC | TACCTCGGCT | TCCCAAAGTG | 16870 |
| 15 | CTGAATTACA | GGCATGAGCC | ACCATGTCTG | GCCTAATTTT | CAACACTTTT | ATATTCTTAT | AGTGTGGGTA | 16940 |
| | TGTCCTGTTA | ACAGCATGTA | GGTGAATTCG | CAATCCAGTC | TGACAGTCGT | TGTTTAACTG | GATAACCTGA | 17010 |
| | TTTATTTTCA | TTTTTTTGTG | ACTAGAGACC | CGCCTGGTGC | ACTCTGATTC | TCCACTTGCC | TGTTGCATGT | 17080 |
| | CCTCGTTCCT | TTGTTTCTCA | CCACCTCTTG | GGTGGCATCT | TGCGTTTCTT | GCCGAGTGTG | TGTTGATCCT | 17150 |
| 20 | CTCGTTGCTT | CCTGGTCACT | GGGCATTTGC | TTTTATTCTT | CTTTGCTTAG | TGTTACCCCT | TGATACCTTG | 17220 |
| | ATTGTCGTTG | TTTGCTTTTG | TTTATTGAGA | CAGTCTCACT | CTGTCAACCA | GGCTGGAGTG | TAATGGCACA | 17290 |
| | ATCTCGGCTC | ACTGCAACCT | CTGCCTCCTC | GGTTCAAGCA | GTTCTCATTC | CTCAACCTCA | TGAGTAGCTG | 17360 |
| | GGATTACAGG | CGCCCCACCAC | CACGCCTGGC | TAATTTTTGT | ATTTTTAGTA | GAGATAGGCT | TTCACCATGT | 17430 |
| | TGGCCAGGCT | GGTCTCAAAAC | TCCTGACCTC | AAGTGATCTG | CCCGCCTTGG | CCTCCCACAG | TGCTGGGATT | 17500 |
| 25 | ACAGGTGCAA | GCCACCGTGC | CCGCGATACC | TTGATCTTTT | AAATGAAGT | CTGAAACATT | GCTACCTTGG | 17570 |
| | TCCTGAGCAA | TAAGACCCCT | AGTGATTTTT | AGCTCTGGCC | ACCCCCCAGC | CTGTGTGCTG | TTTTCCCTGC | 17640 |
| | TGACTTAGTT | CTATCTCAGG | CATCTTGACA | CCCCCACAAG | CTAAGCATT | TAAATATTGT | TTTCCGTGTT | 17710 |
| | GAGTGTTCCT | GTAGCTTTGC | CCCCGCCCTG | CTTTTCTCTC | TTTGTTCCTC | GTCTGTCTTC | TGTCTCAGGC | 17780 |
| | CCGCCGCTGT | GGGTCCCTCT | CCTTGTCTCT | TGCGTGGTTC | TTCTGTCTTG | TTATTGCTGG | TAAACCCAG | 17850 |
| 30 | CTTACCTGTG | GCTGGCCTCC | ATGGCATCTA | GCGACGTCCG | GGGACCTCTG | CTTATGATGC | ACAGATGAAG | 17920 |
| | ATGTGGAGAC | TCACGAGGAG | GGCGGTATC | TTGGCCCGTG | AGTGTCTGGA | GCACCACGTG | GCCAGCGTTC | 17990 |
| | CTTAGCCAGT | GAGTGACAGC | AACGTCCGCT | CGGCCCTGGG | TCAGCCTGGA | AAACCCACAG | CATGTCCGGG | 18060 |
| | TCTGGTGGCT | CCCGCGGTGC | GAGTTTGAAA | TCGCGCAAAAC | CTGCGGTGTG | GCGCCAGCTC | TGACGGTGCT | 18130 |
| | GGCTGGCGGG | GGAGTGTCTG | CTTCTCTCCT | TCTGCTTGGG | AACCAGGACA | AAGGATGAGG | CTCCGAGCCG | 18200 |
| 35 | TTGTGCGCCA | ACAGGAGCAT | GACGTGAGCC | ATGTGGATAA | TTTTAAATTT | TCTAGGCTGG | GCGCGGTGGC | 18270 |
| | TCACGCCTGT | AATCCCAGCA | CTTTGGGAGG | CCAAGCGCGG | TGGATCACGA | GGTCAGGAGG | TCCGAGCAT | 18340 |
| | CTGGCCCAAC | ATGATGAAAC | CCCATCTGTA | CTAAAAACAC | AAAAATTAGC | TGGGCGTGGT | GGCGGCTGCC | 18410 |
| | TGTAATCCCA | GCTACTCGGG | AGGCTGAGCG | AGGAGAATTG | CTTGAACCTG | GGAGTTGGAA | GTTGCAGTGA | 18480 |
| | GCCGACATTG | CACCACTGCA | CTCCAGCCTG | GCAACACAGC | GAGACTCTGT | CTCAAAAAAA | AAAAAAAAAA | 18550 |
| 40 | AAAAAAAATA | AATTCTAGTA | GCCACATTAA | AAAGTGAATA | AAGAAAAAGT | GAAATTATGT | TAAATATAGA | 18620 |
| | TTTTACTGAA | GCCCAGCATG | TCCACACCTC | ATCATTTTAG | GGTGTATTG | GTGGGAGCAT | CACTCACAGG | 18690 |
| | ACATTTGACA | TTTTTTGAGC | TTTGTCTGCG | GGATCCCGTG | TGTAGGTCCC | GTGCGTGGCC | ATCTCGGCTC | 18760 |
| | GGACCTGCTG | GCTTCCCAT | GGCCATGGCT | TTGTATCCAG | ATGGTGCAGG | TCCGGGATGA | GTCGCCAGG | 18830 |
| | CCCTCAGTGA | GCTGGATGTG | CAGTGTCCGG | ATGGTGCACG | TCTGGGATGA | GGTCGCCAGG | CCCTGCTGTG | 18900 |
| 45 | AGCTGGATGT | GTGGTGTCTG | GATGGTGCAG | GTCAGGGGTG | AGGTCTCCAG | GCCTCGGTG | AGCTGGAGGT | 18970 |
| | ATGGAGTCCG | GATGATGCAG | GTCGGGGTGC | AGGTGCGCCG | GCCCTGCTGT | GAGCTGGATG | TGTGGTGTCT | 19040 |
| | GGATGGTGCA | GGTCAGGGGT | AGGTCTCCCA | GGCCCTCGGT | AAGCTGGAGG | TATGGAGTCC | GGATGATGCA | 19110 |
| | GGTCCGGGGT | GAGGTGCGCA | GGCCCTGCTG | TGAGCTGGAT | GTGTGGTGTG | TGGATGGTGC | AGGTCTGGGG | 19180 |
| | TGAGGTCCAC | AGGCCCTGCG | GTGAGCTGGG | TGTGCGGTGT | CTGGATGGTG | CAGGTCTGGA | GTGAGGTCCG | 19250 |
| 50 | CAGACGTGTC | CAGACCATGC | GGTGATGGG | ATATGCGGTG | TCCGGATGGT | GCAGGTCTGG | GGTGAGGTTG | 19320 |
| | CCAGGCCCTG | CTGTGAGTTG | GATGTGGGGT | GTCCGGATGC | TGCAGGTCCG | GTGTGAGGTC | ACCAGGCCCT | 19390 |
| | GCTGTGAGCT | GGATGTGTGG | TGTCTGGATG | GTGCAAGTCT | GGGGTGAAGG | TCGCCAGGGC | CCTGCTTGTG | 19460 |
| | AGCTGGATGT | GTGGTGTCTG | GATGGTGCAG | GTCGTGAGTG | AGGTGCGCCG | GCCCTCGGTG | GCTGGATGTT | 19530 |
| | GCAGTGTCCA | GATGGTGCAG | GTCGGGGTGC | AGGTGCGCCG | ACCCTGCGGT | GAGCTGGATG | TCCGGTGTCT | 19600 |
| | GGATGTGTGA | GGTCTGGAGT | GAGGTGCGCA | GGCCCTCGGT | GAGCTGGATG | TATGGAGTCC | GGATGGTGCC | 19670 |
| 55 | GGTCCGGGGT | GAGGTGCGCA | GACCTGCTG | TGAGCTGGAT | GTGCGGTGTG | TGGATGGTAC | AGGTCTGGAG | 19740 |
| | TGAGGTGCGC | AGACCTGTCT | GTGAGCTGGA | TATGCGGTGT | CCGGATGGTG | CAGGTACAGG | GTGAGGTCTC | 19810 |
| | CAGGCCCTCG | GTGAGCTGGA | GGTATGGAGT | CCGGATGATG | CAGGTCCGGG | GTGAGGTCCG | CAGGCCCTCG | 19880 |
| | TGTGAATCGG | ATGTGCGCGG | TCTGGATGGT | GCAGGTCTGG | GGTGTGGTGC | CCAGGCCCTG | GGTGAAGTGG | 19950 |
| 60 | AGGTATGGAG | TCCGGATGAT | GCAGGTCCGG | GGTGAGGTGC | CCAGGCCCTG | CTGTGAGCTG | GATGTGCGGC | 20020 |
| | GTCTGGATGG | TGCAGGTCTG | GGGTGTGGTC | GCCAGGCCCT | CGGTGAGCTG | GAGGTATGGA | GTCCGGATGA | 20090 |
| | TGCAGGTCCG | GGGTGAGGTT | GCCAGGCCCT | GCTGTGAGCT | GGATGTGCTG | TATCCGGATG | GTGCAAGTCCG | 20160 |
| | GGGTGAGGTC | GCCAGGCCCT | GCTGTGAGCT | GATGTGCTGT | TATCCGGATG | GTGCAAGTCT | GGGTGAGGTT | 20230 |
| | CACCAAGGCC | TGCGGTGAGC | TGGTTGTGCG | GTGTCCGGTT | GCTGCAGGTC | CGGGTGAGT | TGCGCAGGCC | 20300 |
| 65 | CTCGGTGAGC | TGGATGTGCG | GTGTCCCGGT | GTCCGGATGG | TGCAGGTCCA | GGGTGAGGTC | GCTAGGCCCT | 20370 |
| | TGTTGGGCTG | GATGTGCGGT | GTGTGCGCAT | TGCAGGTCTG | GGGTGAGGTC | GCCAGGCCCT | TGGTGGGCTG | 20440 |
| | GATGTGCGGT | GTGTGCGCAT | TGCAGGTCTG | GGGTGAGGTC | GCCAGGCCCT | TGGTGGGCTG | GATGTGTGGT | 20510 |
| | GTCCGGATGG | TGCAGGTCCG | GCGTGAGGTC | GCCAGGCCCT | GCTGTGAGCT | GGATGTGCGG | TGTCTGGATG | 20580 |
| 70 | GTGCAGGTCC | GGGGTGAAGT | AGCCAAGGCC | TTCCGTGAGC | TGGATGTGGG | GTGTCCGGAT | GGTGCAAGTC | 20650 |
| | CGGGGTGAGG | TGCGCAGGCC | CTGCCGCTT | CTGGATATGC | GGTGTCCGGA | TGGTGCAGGT | CCGGGGTGAG | 20720 |
| | GTCAACAGGC | CCTGCCGTTA | GCTGGATGTG | CGGTGTCTGG | ATGGTGCAGG | TCCGGGGTGA | GGTCCGAGG | 20790 |
| | CCCTGCTGTG | AGCTGGATGT | GCTGTATCCG | GATGGTGCAG | GTCGGGGGTG | AGGTCCGCGG | GCCCTGCAGT | 20860 |
| | GAGCTGGATG | TGCTGTATCC | GGATGGTGCA | GGTCTGGCGT | GAGGTCCGCA | GGCCCTGCGG | TTAGCTGGAT | 20930 |
| | ATCGGTGTG | GGATGGTGCA | GGTCCGGGGT | GAGGTACCA | GGCCCTGCGG | TTAGCTGGAT | GTGCGGTGTG | 21000 |
| 75 | CGGATGGTGC | AGGTCTGGGG | TGAGGTCCGC | AGGCCCTGCT | GTGAGCTGGA | TGTGCTGTAT | CCGGATGGTG | 21070 |
| | CAGGTCCGGG | GTGAGGTCCG | CAGGCCCTGC | GGTGAGCTGG | ATGTGCTGTA | TCCGGATGGT | GCAGGTCTGG | 21140 |
| | CGTGAGGTGC | CCAGGCCCTG | CGGTGAGCTG | GATGTGCAGT | GTACGGATGG | TGCAGGTCCG | GGGTGAGGTC | 21210 |
| | GCCAGGCCCT | GCGGTGGGCT | GTATGTGTGT | TGTCTGGATG | GTGCAAGTTC | GGGGTGAAGT | GCCAGGCCCT | 21280 |

| | | | | | | | | |
|----|-------------|-------------|------------|------------|------------|-------------|-------------|-------|
| | TGCGGTGAGC | TGGATGTGTG | GTGTCTGGAT | GCTGCAGGTC | CGGGGTGAGT | TCGCCAGGCC | CTCGGTGAGC | 21350 |
| | TGGATATGCG | GTGTCCCGT | GTCCGAATGG | TGCAGGTCCA | GGGTGAGGTC | GCCAGGCCCT | TGGTGGGCTG | 21420 |
| | GATGTGCCGT | GTCCGGATGG | TGCAGGTCTG | GGGTGAGGTC | GCCAGGCCCT | TGGTGGGCTG | GATGTGGCGT | 21490 |
| 5 | GTCCGGATGG | TGCAGGTCCG | GGGTGAGGTC | ACCAGGCCCT | CGGTGATCTG | GATGTGGCAT | GTCCTTCTCG | 21560 |
| | TTTAAGGGGT | TGGCTGTGTT | CCGGCCGAG | AGCACCGTCT | GCGTGAGGAG | ATCCTGGCCA | AGTTCTCTGA | 21630 |
| | CTGGCTGATG | AGTGTGTACG | TCGTGAGGCT | GCTCAGGTCT | TTCTTTTATG | TCACGGAGAC | CACGTTTCAA | 21700 |
| | AAGAACAGGC | TCTTTTCTA | CCGGAAGAGT | GTCTGGAGCA | AGTTGCAAA | CATTGGAATC | AGGTACTGTA | 21770 |
| | TCCCCACGCC | AGGCCCTGTC | TTCTCGAAGT | CCTGGAACAC | CAGCCCCGCC | TCAGCATGCG | CCTGTCTCCA | 21840 |
| 10 | CTTGCCCTGTG | CTTCCCTGGC | TGTGCAGCTC | TGGGCTGGGA | GCCAGGGGCC | CCGTACACAG | CCTGGTCCAA | 21910 |
| | GTGGATTCTG | TGCAAGGCTC | TGACTGCCTG | GAGCTCACGT | TCTCTTACTT | GTAATAATCAG | GAGTTTGTGC | 21980 |
| | CAAGTGGTCT | CTAGGGTTTG | TAAAGCAGAA | GGGATTTAAA | TTAGATGGAA | ACACTACCAC | TAGCCTCCTT | 22050 |
| | GCCTTTCCCT | GGGATGTGGG | TCTGATTCTC | TCTCTCTTTT | TTTTTCTTTT | TTTGAGATGG | AGTCTCACTC | 22120 |
| | TGTGCCCAG | GCTGGAGTGC | AGTGGCATAA | TCTTGGCTCA | CTGCAACCTC | CACCTCCTGG | GTTTAAGCGA | 22190 |
| 15 | TTACCAGGCC | TCAGCCTCCT | AAGTAGCTGG | GATTACAGGC | ACCTGCCACC | ACGCTGGCT | AATTTTGTGA | 22260 |
| | CTTTTAGGAG | AGACGGGGTT | TCACCATGTT | GGCCAGGGTG | GTCTCGAACT | CATGACCTCA | GGTGATCCAC | 22330 |
| | CCACCTTGGC | CTCCCAAAGT | GCTGGGTTTA | CAGGCTAAGC | CACCGTGCCC | AGCCCCGAT | TCTCTTTTAA | 22400 |
| | TTTCATGCTGT | TCTGTATGAA | TCTTCAATCT | ATTGGATTTA | GGTCATGAGA | GGATAAAATC | CCAGCCACTT | 22470 |
| | GGCGACTCAC | TGACGGGAGC | ACCTGTGCAG | GGAGCACTCG | GGGATAGGAG | AGTTCCACCA | TGAGCTAATC | 22540 |
| 20 | TCTAGGTGGC | TGCATTTGAA | TGGCTGTGAG | ATTTTCTCTG | CAATGTTCGG | CTGATGAGAG | TGTGAGATTG | 22610 |
| | TGACAGATTG | AAGCTGGATT | TGCATCAGTG | AGGAGCAGGA | GCGCTGGTCT | GGGAGATGCC | AGCCTGGCTG | 22680 |
| | AGCCCAGGCC | ATGGTATTAG | CTTCTCCGTG | TCCCGCCAGC | GCTGACTGTG | GAGGGCTTTA | GTCAGAAGAT | 22750 |
| | CAGGGCTTCC | CAGCCTCCCC | TGCACACTCG | AGTCCCTGGG | GGGCTTGTG | ACACCCCATG | CCCCAAATCA | 22820 |
| | GGATGTCTGC | AGAGGGAGCT | GGCAGCAGAC | CTCGTCAGAG | GTAACACAGC | CTCTGGGCTG | GGGACCCCGA | 22890 |
| 25 | CGTGGTGCTG | GGGCCATTTC | CTTGCACTCT | GGGGAGGGTC | AGGGCTTTCC | CTGTGGGAAC | AAGTAAATAC | 22960 |
| | ACAATGCACC | TACTTAGAC | TTTACACGTA | TTTAAATGGT | TGCGACCCAA | CATGGTCATT | TGACCAATAT | 23030 |
| | TTTGAAAGA | ATTTAATTGG | GGTGACCGGA | AGGAGCAGAC | AGACGTGGTG | GTCCCCAAGA | TGCTCCTTGT | 23100 |
| | CACTACTGGG | ACTGTTGTTT | TGCCCTGGGG | GCCTTGAGAG | CCCCCTCTCC | CTGGACAGGG | TACCGTGCTT | 23170 |
| | TTTCTACTGT | GCTGGGCCCT | CGGCTGCGG | TCAGGGCACC | AGCTCCGGAG | CACCCGCGCG | CCACGTGCTC | 23240 |
| 30 | ACGGAGTGCC | AGGCTGTGAG | CCACAGATGC | CCAGGTCCAG | GTGTGGCCGC | TCCAGCCCCC | GTGCCCCCAT | 23310 |
| | GGGTGTTT | GGGGGAAAAG | GCCAAAGGGA | GAGGTGTGAG | GAGACTGGTG | GGCTCATGAG | AGCTGATTCT | 23380 |
| | GCTCCTTGGC | TGAGCTGCC | TGAGCAGCCT | TCCCCGCCCT | CTCCATCTGA | AGGGATGTGG | CTCTTTCTAC | 23450 |
| | CTGGGGTCC | TGCCCTGGGG | CAGCCTTGGG | CTACCCAGT | GGCTGTACCA | GAGGGACAGG | CATCCTGTGT | 23520 |
| | GGAGGGGAC | GGGTTACAGT | GGCCCCAGAT | GCAGCCTGGG | ACCAGGCTCC | CTGGTGCTGA | TGGTGGGACA | 23590 |
| 35 | GTACCCCTGG | GGGTTGACCG | CCGAGCTGGG | CGTCCCCAGG | GTTGACTATA | GGACCAGGTG | TCCAGGTGAC | 23660 |
| | CTGCAAGTAG | AGGGGCTCTC | AGAGGCTGCT | GCGTGGCATG | GGTGAGACGT | GCCCCGGGGA | TGGCTTCCCG | 23730 |
| | CGTGTGCTGC | CGTGGGTGCC | CTGAGCCCTC | ACTGAGTCGG | TGGGGGCTTG | TGGCTTCCCG | TGAGCTTCCC | 23800 |
| | CCTAGTCTGT | TGCTGTGGCT | AGCAAGCCTC | CTGAGGGGCT | CTCTATTGCA | GACAGCACTT | GAAGAGGGTG | 23870 |
| 40 | CAGTTCGCGG | AGCTTCGGA | AGCAGAGGTC | AGGCAGCATC | GGGAAGCCAG | GCCCCGCCCT | GTGACGTCCA | 23940 |
| | GACTCCGCTT | CATCCCCAAG | CCTGACGGGC | TGCGGCCGAT | TGTGAACATG | GACTACGTCG | TGGGAGCCAG | 24010 |
| | AACCTTCCGC | AGAGAAAAGA | GGGTGGCTGT | GCTTTGTGTT | AACTTCTCTT | TTAAACAGAA | GTGCGTTTGA | 24080 |
| | GGCCCACTT | TGGTATCAGC | TTAGATGAAG | GGCCCGGAGG | AGGGGCCACG | GGACACAGCC | AGGGCCTATG | 24150 |
| | CACGGCGCCA | ACCCATTGTT | GCGCACAGTG | AGGTGGCCGA | GGTGCCGGTG | CCTCCAGAAA | AGCAGCGTGG | 24220 |
| 45 | GGGTGTAGGG | GGAGCTCCTG | GGGCAGGGAC | AGGCTCTGAG | GACCAACAAG | AGCAGCCGGG | CCAGGGGCTG | 24290 |
| | GATGCAGCAC | GGCCCGAGGT | CCTGGATCCG | TGTCCTGCTG | TGGTGCCGAG | CCTCCGTGCG | CTTCCGCTTA | 24360 |
| | CGGGGCCCCG | GGACCAGGCC | ACGACTGCCA | GAGGCCACCC | GGGCTCTGAG | GATCCTGGAC | CTTGCCCCAC | 24430 |
| | GGCTCCTGCA | CCCCACCCCT | GTGCTGCGG | TGGCTGCGGT | GACCCCGTCA | TCTGAGGAGA | GTGTGGGGTG | 24500 |
| 50 | AGGTGGACAG | AGGTGTGGCA | TGAGGATCCC | GTGTGCAACA | CACATGCGGC | CAGGAACCCG | TTTCAAACAG | 24570 |
| | GGTCTGAGGA | AGCTGGGAGG | GGTCTAGGTT | CCCGGCTCTG | GGTGCTGGG | GACACTGGGG | AGGGGCTGCT | 24640 |
| | TCTCCCTTGG | GTCCCTATGG | TGGGGTGGGC | ACTTGGCCGG | ATCCACTTTC | CTGACTGTCT | CCCATGCTGT | 24710 |
| | CCCCGCCAGG | CCGAGCGTCT | CACCTCGAGG | GTGAAGGCAC | TGTTTACGCT | GCTCAACTAC | GAGCGGGGCG | 24780 |
| | GGCGCCCCGG | CCTCCTGGGC | GCCTCTGTGC | TGGGCTGGGA | CGATATCCAC | AGGGCCTGCG | GCACCTTGTG | 24850 |
| | GCTGCGTGTG | CGGGCCAGG | ACCCGCCGCC | TGAGCTGTAC | TTTGTCAAGG | TGGGTGCCGG | GGACCCCGCT | 24920 |
| | GAGCAGCCCT | GCTGGACCTT | GGGAGTGGCT | GCCTGATTGG | CACCTCATGT | TGGGTGGAGG | AGGTACTCCT | 24990 |
| 55 | GGGTGGGCGG | CAGGGAGTGC | AGGTGACCTT | GTCACTGTTG | AGGACACACC | TGGCACCTAG | GGTGGAGGCC | 25060 |
| | TTACGCCCTT | CCTGCAGCAC | ATGGGGCCGA | CTGTGCCACC | TGACTGCCCC | GGCTCCTATT | CCCAAGGAGG | 25130 |
| | GTCCCACTGG | ATTCCAGTTT | CCGTCAGAGA | AGGAACCGCA | ACGGCTCAGC | CACCAAGGCC | CGGTGCCCTT | 25200 |
| | CACCCAGTC | CTGAGCCAGG | GGTCTCCTGT | CCTGAGGCTC | AGAGAGGGGA | CACAGCCCGC | CCTGCCCTTG | 25270 |
| 60 | GGGTCTGGAG | TGTTGGGGGT | CAGAGAGAGA | GTGGGGGACA | CCGCCAGGCC | AGGCCCTGAG | GGCAGAGGTG | 25340 |
| | ATGTCTGAGT | TTCTGCGTGG | CCACTGTGAG | TCTCCTCGCC | TCCACTCACA | CAGGTGGATG | TGACGGGCGC | 25410 |
| | GTACGACACC | ATCCCCCAGG | ACAGGCTCAC | GGAGGTTCAT | GCCAGCATCA | TCAAACCCCA | GAACACGTAT | 25480 |
| | TGCGTGGCTC | GGTATGCCGT | GGTCCAGAAG | GCCGCCCATG | GGCACGTCCG | CAAGGCCCTT | AAGAGCCACG | 25550 |
| | TAAGGTTTAC | GTGTGATAGT | CGTGTCCAGG | ATGTGTGTCT | CTGGGATATG | AATGTGTCTA | GAATGCAGTC | 25620 |
| 65 | GTGTCTGTGA | TGCGTTTCTG | TGGTGGAGGT | ACTTCCATGA | TTTACACATC | TGTGATATGC | GTGTGTGGCA | 25690 |
| | CGTGTGTGTC | TGTTGTGCATG | TATCTGTGGC | TGTCATATT | GTGGTGTGTG | TGTGTGTGGC | ACGTTGTGTG | 25760 |
| | CCATGGTGTG | TGTGCCTGTG | GTGTGCATGT | GTGTGTGTCT | GTGACACGTG | CATGTTTCATG | CTGTGTGCTG | 25830 |
| | CATGTCTGTG | ATGTGGCTAT | TTGTGGTGTG | TGTGTGCATG | TGTCCGTGAC | ATATGCGTGT | CTATGGCATG | 25900 |
| | GGTGTGTGTG | GCCCTTGGC | CTTACTCCTT | CCTCCTCCAG | GCATGGTCCG | CACCATTTGC | CTCACGCTCT | 25970 |
| | CGGGTGCTGG | TTTGGGGAGC | TCCACATTCA | GGGTCTCTAC | TTCTAGCATG | GGTGCCCTG | TCCGTCTACA | 26040 |
| 70 | GGGCTGGGCC | TTGGAGACTG | TAAGCCAGGT | TTGAGAGGAG | AGTAGGGATG | CTGGTGGTAC | CTTCTGGGAC | 26110 |
| | CCCTGGCACC | CCCAGGACCC | CAGTCTGGCC | TATGCCGGCT | CCATGAGATA | TAGGAAGGCT | GATTACAGGC | 26180 |
| | TGCGTCCCCG | GGACACACTC | CTCCAGAGC | GGCCGGGGGC | CTTGGGGCTC | GGCAGGGGTG | AAAGGGGGCC | 26250 |
| | TGGGCTTGGG | TTCCACCCCA | GTGGTCATGA | GCACGCTGGA | GGGGTAAGCC | CTCAAAGTCG | TGCCAGGGCCG | 26320 |
| | GGGTGCAGAG | GTGAAGAAGT | ATCCCTGGAG | CTTCGGTCTG | GGGAGAGGCA | CATGTGGAAA | CCCACAAGGA | 26390 |
| 75 | CCTCTTTCTC | TGACTTCTTG | AGCT | | | | | 26414 |

Contig 2:

| | | | | | | | | |
|----|-------------|-------------|------------|-------------|-------------|-------------|------------|------|
| | TGTGGGATTG | GTTCATGT | GTGGGATAGG | TGGGGATCTG | TGGGATTGGT | TTTTATGAGT | GGGGTAACAC | 70 |
| | AGAGTTCAAG | GCGAGCTTTC | TTCTGTAGT | GGGTCTGCAG | GTGCTCCAAC | AGCTTTATTG | AGGAGACCAT | 140 |
| 5 | ATCTTCCTTT | GAACATAGGT | CGGGTTTATA | GTAAGTCAGG | GGTGTGGAGG | CCTCCCCCTGG | GCTCCCTGTT | 210 |
| | CTGTTTCTTC | CACCTCTGGG | TCGTGTGGTG | CCTGCTGTGG | TGTGTGGCCG | GTGGGCAGGG | CTTCCAGGCC | 280 |
| | TCCTTGTGTT | CATTGGCCTG | GATGTGGCCC | TGGCTACGCT | CCGTCCTTGG | AATTCCCCCTG | CGAGTTGGAG | 350 |
| | GCTTTCTTTC | TTTTCTTTT | TCCTTCTTTT | TTTTTTTTTT | TGATAACAGA | GTCTCGCTCT | TTTTTGCCCA | 420 |
| | GGCTGGAGTG | GTTCGGCGTG | ATCTTGCGTC | ACTGCAACCT | GTGCTTCCTG | AGTTCAAGCA | ATTCTCTTGC | 490 |
| 10 | CTCAGCCTCC | CAAGTAGCTG | GAATTATAGG | CGCCCACCAC | CATGCTGACT | AATTTTTGTA | ATTTTAGTAG | 560 |
| | AGACGAGGTT | TCTCCATGTT | GGCCAGGCTG | GTCTCGAACT | CCTGACCTCA | GGTGATCCTC | CCACCTCGGC | 630 |
| | CTCCCAAAGT | GCTGGGATGA | CAGGTGTGAA | CCGCCGCGCC | CGGCCGAGAC | TCGCTTCCTG | CAGCTTCCGT | 700 |
| | GAGATCTGCA | GCGATAGCTG | CCTGCAGCCT | TGGTGTCTGAC | AACCTCCGTT | TTCTTCTCTC | AGGTCTCGCT | 770 |
| | AGGGGTCTTT | CCATTTCTATG | ACTCTCTTCA | CAGAAGAGTT | TCACGTGTGC | TGATTTCCCG | GCTGTTTCTT | 840 |
| 15 | GCGTAATTGG | TGCTGTGCTG | TTATCGATGG | CCTCCTTCCA | TTTCCTTTAG | GCTTTGTTTA | TGTGTGTTTT | 910 |
| | TCCGGCTCCT | TGAAGGAAAA | GTTCGATTA | TGGATGTGTTG | AACCTTCTTT | TCTAAACAAG | CATCTGAAGT | 980 |
| | TGCCGTTTTT | CCTCTAAAGC | AGGGATCCCG | AGGCCCTGG | CTGTGGAGTG | GCACCGGTCT | GGGGCTGTT | 1050 |
| | AGGAACCCGG | GCGACAGCGG | GAGGCTAGGT | GGGGTGTGGG | GAGCCAGCGT | TCCCGCCTGA | GGCCCCCCCC | 1120 |
| 20 | TCTCAGATCA | GCAGTGGCAT | GCGGTGCTCA | GAGGCCACAC | CACCCTACTG | AGAACTGTGT | GTGAGAGGGG | 1190 |
| | TCTAGATTCT | GTGCTCCTTA | TGGGAATCTA | ATGCTGATG | ATCTGAGGTG | GAACCGTTTG | CTCCCAAAAC | 1260 |
| | CATCCCTTTC | CCCCTGCTG | TCCTGTGGAA | AAATCGTCTT | CCACGAAACC | AGTCCCTGGT | ACCACAATGG | 1330 |
| | TTGGGGACCC | TGTGCTAAAG | ACCTGCTTCA | GCAGCCTCTC | GTGAGTGTG | ATATATTGGC | TTTTCTGTGT | 1400 |
| | TGAGTCCAGA | ATAATTACGG | ATTCTGTGA | TGCTTTCCCG | CGACCTCAGA | CCCATGGGCT | ATTTGTGGGC | 1470 |
| 25 | GTGTTGCGCTG | CTCCTGGGTT | GGGAAGGGTG | CAGGCCCAT | GTACCTTCCT | GTTACTGCCT | TCCAGGTTGG | 1540 |
| | TTCTCAGGGT | TGAATCGTAC | TEGATGTGGT | TTTAGCCAC | GGCCCTGCGG | CCAGCTCCTG | GGGGCTGGGG | 1610 |
| | AACATGCTGA | AGCACAGAGT | CACCGTCCGC | GTCTTTTGAT | GCCTCACAA | CTCGAGGCTT | CCTGTCTCCG | 1680 |
| | TGTTAGTGTG | TGTCACGTGC | CTGCTCACAT | CCTGTCTTGG | GGACGCAGGG | GCTTAGCAGG | TCCCGTAGTA | 1750 |
| | AATGACAAGC | GTCTTGGGGG | AGTCTGCAGA | ATAGGAGGTG | GGGGTGCCGG | TCTCTCTCCC | GGGTCTTCAG | 1820 |
| 30 | ACTCTTCTCC | TGCCGTGTCT | GTGGCTGCAC | CTGCATCCCT | GCAATCCCTC | CAGCACTGGG | CTGGAGAGGC | 1890 |
| | CCGGGAGCTC | GAGTGCCACT | TGTGCCACGT | GACTGTGGAT | GGCAGTCGGT | CACGGGGGTC | TGATGTGTGG | 1960 |
| | TGACTGTGGA | TGGCGGTTGG | TCACAGGGGT | CTGATGTGTG | GTGACTGTGG | ATGGCGGTG | TGGGGTCTGA | 2030 |
| | TGTGTGTGAGT | GTGGATGGCG | GTGCTGGGTT | CTGATGTGTG | GTGACTGTGG | ATGGCGGTG | TGGGGTCTGA | 2100 |
| | TGTGTGTGACT | GTGGATGGCG | GTGCTGGGTT | CTGATGTGTG | GTGACTGTGG | ATGGCGGTG | TGGGGTCTGA | 2170 |
| 35 | TGGTGTGACT | GGATGGCAGT | CGTGGGGTCT | GATGTGTGGT | GACTGTGGAT | GGCGGTCTGT | GGGTCTGATG | 2240 |
| | TGGTGTGACT | GGATGGCAGT | CGTGGGGTCT | GATGTGTGGT | GACTGTGGAT | GGCGGTCTGT | GGGTCTGATG | 2310 |
| | TGTGTGTGACT | GTGGATGGCG | GTGCTGGGTT | CTGATGTGTG | GTGACTGTGG | ATGGCGGTG | TGGGGTCTGA | 2380 |
| | TGTGTGTGTA | CTGTGGATGG | CGTCTGGTGG | GTCTGTGTG | GTGACTGTGG | ATGGCGGTG | TGGGGTCTGA | 2450 |
| | TGTGTGTGTA | CTGTGGATGG | TGATCGGTGA | CAGGGGTCTG | ATGTGTGGTG | ACTGTGGATG | GCGGTCTGGG | 2520 |
| 40 | GGTCTGAGT | TGGGTGACTG | TGGATGGTGA | TCGGTTCACAG | GGGTCTGATG | TGTGGTGA | TGGGTCTGGG | 2590 |
| | GTGCTGGGTT | CTGATGTGTG | GTGACTGTGG | ATGGCGGTG | GTCCCGGGGG | TCTGATGTGT | GGTACTGTGT | 2660 |
| | GATGGCGATC | GGTCACAGGG | GTCTGATGTG | TGGTGTGACT | GGATGGCGGT | CGTGGGGTCT | GATGTGTGGT | 2730 |
| | GACTGTGGAT | GGCGGTCTGT | GGGTCTGATG | TGGTGTGACT | GGATGGCGGT | CGTGGGGTCT | GATGTGTGGT | 2800 |
| | GACTGTGGAT | GGCGGTCTGT | GGGTCTGATG | TGGTGTGACT | GGATGGCGGT | CGTGGGGTCT | GATGTGTGGT | 2870 |
| 45 | GACTGTGGAT | GGCGGTCTGT | GGGTCTGATG | TGGTGTGACT | GGATGGCGGT | CGTGGGGTCT | GATGTGTGGT | 2940 |
| | GTGGTGTGACT | TGGATGGCAG | TCGTGGGGTG | TGATGTGTGG | TGACTGTGGA | TGGCGGTCTG | GGGGTCTGAT | 3010 |
| | GTGTTGTGAC | TGTGGATGGC | GGTCTGGGGG | TCTGATGTGT | GGTACTGTG | GATGGCGGTC | GTGGGGTCTG | 3080 |
| | ATGTGTGGTG | ACTGTGGATG | GCGGTCTGGG | GGTCTGATGT | GGTACTGTG | GATGGCGGTC | GTGGGGTCTG | 3150 |
| | ATGTGTGGTG | ACTGTGGATG | GTGATCGGTC | ACAGGGGTCT | GATGTGTGGT | GACTGTGGAT | GGCGGTCTGT | 3220 |
| 50 | GGGTCTGATG | TGTGGTACT | GTGGATGGCG | GTGCTGGGTT | CTGATGTGGT | GACTGTGGAT | GGCGGTCTGT | 3290 |
| | GGGTCTGATG | TGTGGTACT | GTGGATGGCG | GTGCTGGGTT | CTGATGTGGT | GACTGTGGAT | GGCGGTCTGT | 3360 |
| | GTCACAGGGG | TCTGATGTGT | GGTACTGTG | GATGGCGGTC | GTGGGGTCTG | ATGTGTGGTG | ACTGTGGATG | 3430 |
| | GCGGTCTGTT | GGTCTGATGT | GTGGTACTG | TGGATGGCGG | TCGTGGGGTG | TGATGTGTGG | TGACTGTGGA | 3500 |
| | TGGCGGTCTG | GGGGTCTGAT | GTGGTACTG | TGGATGGTGA | TCGGTACACAG | GGGTCTGATG | TGTGGTAGCT | 3570 |
| 55 | GCAGGTGGAG | TCCCAAGGTG | GTCTGTAGCT | ACTTTGCGTC | CTCGCCCCCT | CGGCCCCCTG | TTCCCAAAAC | 3640 |
| | GAAGCTTCCC | AGGCGTCTC | TGGGCTTCAT | CCCGCCATCG | GGCTTGGCCG | CAGGTCCACA | CGTCTGTATC | 3710 |
| | GGAAAGAAAC | AGTCCCAAGC | TCTGGCCGGG | CAGGCCCCACA | TTTGTGGCTC | ATGCCCTCTC | CTCTGCCGGC | 3780 |
| | AGGTCTCTAC | CTTGACAGAC | CTCCAGCCGT | ACATGCGACA | GTTCTGTGGT | CACCTGCAGG | AGACCAGCCC | 3850 |
| | GCTGAGGGAT | GGCGTCTGTA | TCGAGCAGGT | CTGGGCACTG | CCCTGCAGGG | TTGGGCACGG | ACTCCAGACA | 3920 |
| 60 | GTGGGTCTCT | CCCTGGGCAA | TCAGTGGGCT | CATGACCGGA | CAGACTGTGT | GGCCTGGGGG | GCAGTGGGGG | 3990 |
| | GAATGAGCTG | TGATGGGGGC | ATGATGAGCT | GTGTGCCTTG | GCGAAATCTG | AGCTGGGGCA | TGCCAGGCTG | 4060 |
| | CGACAGCTGC | TGCATTACAG | CACCTGTCTA | CGTTTGACTG | CGCGGCTCTT | CTCCAGTTCC | GCAGTGCCTT | 4130 |
| | TGTTTATGAT | TTGCTAAATG | TCTTCTCTG | CAGTTTGTAGT | CTTGAGGCCA | AAGGAAAGGT | GTCCCTCTCC | 4200 |
| | TTTAGGAGGG | CAGGCCATGT | TTGAGCCGTG | TCCTGCCACG | CTGGCCCCCT | AGTGTGGGT | CTGAGGCCAA | 4270 |
| 65 | AGGAACAGTG | TCCCTCTTCT | TAGGAGGAGC | GGCGGTGTTT | GAGCCACGCC | CCGCTGAGCG | GGCCTCTCAG | 4340 |
| | TGCTGGGTCT | TGCCACGTGG | CCTGTGGGCC | CTTTGAGAGT | GTGGTCTGTC | CACGTGGGCC | TGTGGCTCTT | 4410 |
| | TGCAGATGCC | TGTTAGCACT | TGCTCGGCTC | TAGGGGACAG | TCGTGTCCAC | CGCATGAGGC | TCAGAGACCT | 4480 |
| | CTGGGCGAAT | TTCTTTGGCT | CCCAGGGTGG | GGGTGGAGGT | GGCCTGGGCT | GCTGGGACCC | AGACCCTGTG | 4550 |
| | CCCCGAGCT | GGGCAGCAAC | TCCTGGATCA | CATATGCCAT | CCGGGCCACG | GTGGGCTGTG | TGGGTGTGAG | 4620 |
| 70 | CCCAGCTTGA | CCCACAGGTG | CCCCAGAGGA | GACGTTCTGT | GTACACACT | CTGGCTAAGC | CCATGTGTGT | 4690 |
| | CTGCAGAGAC | TCGGCCCGGC | CAGCCACAGA | TGGCCCTGCA | TTCCAGCCCA | GCCCCGCACT | TCATCACAAA | 4760 |
| | CACTGACCCC | AAAAGGGACG | GAGGGTCTTG | GCCACGTGGT | CCTGCCTGTC | TCAGCACCCA | CCGGCTCACT | 4830 |
| | CCCATGTGTC | TTCCGCTCTG | TTTCGACAGG | CTCCTCCCTG | AATGAGGCCA | GCAGTGGGCT | CTTCGACGTC | 4900 |
| | TTCTACGCT | TCATGTGCCA | CCACGCCGTG | CGCATCAGGG | GCAAGTGAGT | CAGGTGGCCA | GGTGCCATTG | 4970 |
| | CCCTGCGGGT | GGCTGGGCGG | GCTGGCAGGG | CTTCTGTCTA | CCTCTCTCTT | GCCCCCTCCC | CACTGNCCTT | 5040 |

| | | | | | | | | |
|----|------------|-------------|------------|------------|-------------|------------|------------|-------|
| | CTGCCCGGGG | CCACCAGAGT | CTCCTTTTCT | GGCCCCCGCC | CCCTCCGGCT | CCTGGGCTGC | AGGCTCCCGA | 5110 |
| | GGCCCCGGAA | ACATGGCTCG | GCTTGCGGCA | GCCCGAGCGG | AGCAGGTGCC | ACACGAGGCC | TGGAAATGGC | 5180 |
| | AAGCGGGGTG | TGGAGTTGCT | CCTGCGTGGA | GGACGAGGGG | CGGGGGGTGT | GTCTGGGTCA | GGTGTGCGCC | 5250 |
| 5 | GAGCGTTTGA | GCCTGCAGCT | TGTGAGCTCC | AAGTTACTAC | TGACGCTGGA | CACCCGGCTC | TCACACGCTT | 5320 |
| | GTATCTCTCT | CTCCCGATAC | AAAAGGATTT | TATCCGATTG | TCATTCTCTG | CCCTGTCTGT | TGACCCCGCG | 5390 |
| | GAGGGCGGCG | GCTCTTCTCT | CTGTGACTAG | ATTCCCATC | TGGAAGTGC | GGGGTTGACC | GTGTAGTTTG | 5460 |
| | CTCCTCTCGG | GGGGCTGTGT | TGTGGCATGG | GGCAGGCGGC | CTGGGAGAGC | TGCCGTACAC | CAGCCACTGG | 5530 |
| | GTGAGCCACA | CTCAGCGTGG | TAGAGCCACA | GTGCCTGGTG | CCACATCAGC | TCCTCTGGAT | TTTAAAGTAA | 5600 |
| 10 | ACCACACACC | TCCCGGCAGG | CATCTGCCTG | CGACCCCTGT | TGTGCCTGGG | GAGAGTGGTA | GCACGGAGGA | 5670 |
| | AATTCGTGCA | CACTCAAGGT | CATCAGCAAG | GTCATCCGCA | GTCAGGTGGA | ACGTGGAGGC | CTCTCTCTGG | 5740 |
| | GATCGTCTCC | AGCGGATAAA | GGACTGTGCA | CAGCTTCGGA | AGCTTTTATT | TAAAAATATA | ACTATTAAAT | 5810 |
| | ATTGCATTAT | AAGTAATCAC | TAATGGTATC | AGCAATTATA | ATATTTATTA | AAGTATAATT | AGAAATATTA | 5880 |
| | AGTAGTACAC | ACGTTCTGGA | AAAACACAAA | TTGCACATGG | CAGCAGAGTG | AATTTTGGCC | GAGGGACACG | 5950 |
| 15 | TGTGCACATG | TGTGTAAGCG | GCCCCCAGGC | CCACAGAAAT | CGCTGACAAA | GTCACCTCCC | CAGAGAAGCC | 6020 |
| | ACCACGGGGC | TCCTTCGTGG | TCGTGAATTT | TATTAAGATG | GATCAAGTCA | CGTACCGTCC | ACGTGTGGCA | 6090 |
| | GGGCTTTGGG | GAAATGTGAGG | TGATGACTGC | GTCTCATGCG | CCTGACAGAC | AGGAGGTGAC | TGTGCTGTGC | 6160 |
| | CTGTCCCTAG | GACACGGACA | GGCCCGAAGC | TCTAGTCCCC | ATCGTGGTCC | AGTTTGGCCT | CTGAATAAAA | 6230 |
| | ACGTCTTCAA | AACCTGTTGC | CCCAAAAAC | AAGAACAGAG | AGAGTTTCCC | ATCCCATGTG | CTCACAGGGG | 6300 |
| 20 | CGTATCTGCT | TGCGTTGACT | CGCTGGGCTG | GCCGAGCTCC | TAGAGTTGGT | CGGTGTGCTT | CTGTGCTGCT | 6370 |
| | AGTGCAGTCC | TCTTGCCCAT | CACTGTGATA | TCTGCACCAG | CAAGGAAAGC | CTCTTTTCTT | TTCTTTTCTT | 6440 |
| | TTTTTTTTTT | GAGACGGAAC | GTCAGTGTG | TCTGCCTGGG | CTTGAGTGCA | GTGGCGCGAT | CTCAACTCAC | 6510 |
| | TGCAACCTCC | GCCCTCCGGG | TTCCAGCATT | TCTCTGCTCG | CAGCCTCCCG | AGCAGCTGAG | ATTACAGGCA | 6580 |
| | CCCACCCCTC | GCGCCTGGCT | AATTTTTGTA | TTTTTAGTAG | AGAGGGGTTT | TTGCCATGTT | GGCCAGGCTG | 6650 |
| 25 | GTCTCGAATT | CCTGACCTCA | GGTGATCCAC | CCACCTCGGC | CTCCCAAAGT | GCTGGGATTA | CAGGTGTGAG | 6720 |
| | CCATCACGCC | CAGCCGGAAA | GCCCTCTTTT | AAGGTGACCA | CCTATAGCGC | TTCCCGAAAA | TAACAGGTCT | 6790 |
| | TGTTTTTGCA | TAGAGGTGCA | AGCGTCTCTT | AGCAACAGGA | GTGGCGTCTT | GTGGGCTCTG | GGGATGCTGT | 6860 |
| | AGGGTTCGGT | GGCAGCCATG | CCTTCTGTGT | GCACCTTAGG | GTTCCACGGG | GCTATTCTGC | TCTCACTGTT | 6930 |
| 30 | TGTCTGAAAA | CGACCCCTTG | GCATCCTTTG | TTGGAGAGTT | TCTGCTTCTC | GTTGGTCATG | CTGAAACTAG | 7000 |
| | GGGCAAGGTT | TATCCGTTG | GCGCGCAGCG | GCTACATGTA | GGGTCAATGAG | TCTTTCACCG | TGGACAAATT | 7070 |
| | CCTTGAAAAA | AAAAAAGGA | GTCCGGTTAA | GCATTCTATC | CGGGTCAAGT | GTCTGGTTCT | GTGAATAAAC | 7140 |
| | TCTAAGATTG | AAGAAACCTT | AATGAAAGAA | AACCTTGATG | ATTCAGAGCA | AGGATGTGGT | CACACCTGTG | 7210 |
| | GCTGTGATCT | TTTTAGCCGC | CCCAGTGAT | GGTGAGAGTG | GGGAGCAGGG | ATTGTTTGTG | CAGAGGTCTC | 7280 |
| | ATCTGGTATG | TTTCTGAGGT | GTTTGCCGGC | TGAATGGTAG | ACGTGTGCTT | TGTGTGTATG | AGGTTCTGTG | 7350 |
| 35 | TCTGTGTGTG | GCTCGGTTTG | AGTGATACGA | GTGCCAGCAC | ATGCCCTGCC | CGTCTCTCAC | CTGTGTCTTC | 7420 |
| | CCGCCCCAGG | TCTTACGTCC | AGTGCCAGGG | GATCCCGCAG | GGCTCCATCC | TCTCCACGCT | GCTCTGCAGC | 7490 |
| | CTGTGCTACG | GCGACATGGA | GAACAAGCTG | TTTGCGGGGA | TTGCGCGGGA | CGGGTGAGGC | CTCCTCTTCC | 7560 |
| | CCAGGGGGGG | TTGGGTGGGG | GTTGATTTGC | TTTTGATGCA | TTTCACTGTA | ATATTCCTGG | TGCTCTGGAG | 7630 |
| 40 | ACCATGACTG | CTCTGTCTTG | AGGAACCGA | CAAGGTTGCA | GCCCCCTCTT | GGTATGAAGC | CGCACGGGAG | 7700 |
| | GGGTTGCACA | GCTTGAGGAC | TGCGGCTCTC | ACGCAGGCTC | TGTCCAGCGG | CCATGTCCAG | AGGCCCTCAG | 7770 |
| | GCTCAGCAGG | CGGGAGGGCC | GCTGCCCTGC | ATGATGAGCA | TGTGAATTCA | ACACCGAGGA | AGCACACCA | 7840 |
| | CTTCTGTCTC | GTCACCCAGG | TTCCGTTAGG | GTCCTTGGGG | AGATGGGGCT | GGTGACGCTT | GAGGCCCCAC | 7910 |
| | ATCTCCAGCT | AGGCCCTCGA | CAGGTGGGCT | GGACTGGGCG | CCTCTTCAGC | CCATTGCCCA | TCCCACTTGC | 7980 |
| | ATGGGGTCTA | CACCCAAAGG | CGCACACACC | TAAATATCTG | GCCAACTTAA | TGTGGTTCAA | CTCAGCTGGC | 8050 |
| 45 | TTTTATTGAC | AGCAGTTACT | TTTTTTTTTT | TAATACTTTA | AGTTCTAGGG | TACATGTGCA | CGACGTGCAG | 8120 |
| | GTAGTTTACA | TAGTATATACA | TGTGCCATGT | TGGTGTGCTG | CACCCATTAA | CTCATCATTT | ACATTAGGTA | 8190 |
| | TATCTCTCTA | TGCTATCCCT | CCCCACTCCC | CCCATCCCAT | GACAGGCCCT | GGTGTGTGAT | GTTCCCCACC | 8260 |
| | CTGTGTCCAA | GTTTCTCTAT | TGTTCAAGTT | CCACCTGTGA | GTGAGAACAT | GTGGTGTGTT | GTTTCTCTTC | 8330 |
| | CTTGCAATAG | TTTGCTCAGA | GTGATGGTTT | CCAGCTTCGT | ACATGTCCCT | ACAAAGGACA | TGAACCTATC | 8400 |
| 50 | CTTTTTTATG | ACTGCATAGT | ATTCCGTGGT | GTATATGTGC | CACATTTTCT | TAATCCAGTC | TATCATCGAT | 8470 |
| | GGACATTGCG | GTTGGTTGCA | AGTCTTTGCT | ACTGTGAATA | GTGCCGCAAT | AAACATACGT | GTGCAATGAT | 8540 |
| | CTTTATAGCA | GCATGATTTA | TAATCCTTTG | GGTATATACC | CAGTAATGGG | ATGGCTGGGT | CAAAATGGAT | 8610 |
| | TTCTAGTTCT | AGATCCTTGA | GGAATCAGCA | CACCTGTCTC | CACAAATGGT | GAACCTAGTT | ACACCTCCAC | 8680 |
| | CAACAGTGTA | AAAGTGTCT | GGTGCTGGAG | AGGATGTGGA | CAGCAGTTAT | TTTTTTATGA | AAATAGTATC | 8750 |
| 55 | ACTGAACAAG | CAGACAGTTA | GTGAAGGATG | CGTCAGGAAG | CCTGCAGGCC | ACACAGCCAT | TTCTCTCGAA | 8820 |
| | GACTCCGGGT | TTTTCTGTGT | CATCTTTTGA | AACTCTAGCT | CCAATTATAG | CATGTACAGT | GGATCAAGGT | 8890 |
| | TCTTCTTCAT | TAAGGTTCAA | GTTCTAGATT | GAAATAAGTT | TATGTAACAG | AAACAAAAAT | TTCTTGATCA | 8960 |
| | CACAACCTGC | TCTGGGATTT | GGAGGAAAGT | GTCTCTGAGC | TGGCGGCACA | CTGGTCAGCC | CTCTGGGACA | 9030 |
| | GGATACTCTC | GGCCCATGGT | CATGGGGGCT | TGGCTTGGG | CCTGAGGGTC | ACACAGTGCA | CCATGCCACG | 9100 |
| 60 | CTTCTGTGG | ATAGGATCTG | GGTCTCGGAT | CATGCTGAGG | ACCACAGCTG | CCATGCTGGT | AAAGGGCACC | 9170 |
| | ACGTGGCTCA | GAGGGGGCGA | GGTTCCACAG | CCCAGCTTTC | TTACCGTCTT | CAGTTATTTT | TCCCTAAGAG | 9240 |
| | TCTGAGAAGT | GGGGCCGCGC | CTGATGGCCT | TCGTTCTGCT | TCAGCTGGCA | CAGAATTGCA | CAAGCTGATG | 9310 |
| | GTAAACACTG | AGTACTTATA | ATGAATGAGG | AATTGCTGTA | GCAGTTAACT | GTAGAGAGCT | CGTCTGTTGG | 9380 |
| | AAAGAAATTT | AAGTTTTCAT | TTTAAACCGT | TTGGAGAAAT | TTACTTTATT | TATGGCTGTG | TAAATGTTT | 9450 |
| 65 | GACATTCAGT | CCCTCGTAGA | CAGATACTAC | GTAAGAAAGT | TAAAGTTAAT | CTTGCTGTGT | ATTTTCCCTT | 9520 |
| | ATTTTAGGCT | GCTCTGCGT | TTGGTGGATG | ATTCTTGTGT | GGTGACACCT | CACCTCACCC | ACGCGAAATC | 9590 |
| | CTTCTCTCAG | TGAGGCCCCG | GCCGTGTGTC | TGTGGGGACC | TCCACAGCCT | GTGGGCTTTG | CAGTTGAGCC | 9660 |
| | CCCCGTGTCC | TGCCCTGTGC | ACCGCAGCGT | TGTCTCTGCC | AAGTCTCTCT | TCTCTGCCGG | TGCTGGATCC | 9730 |
| | GCAAGAGCAG | AGGCGCTTGG | CCGTCACACC | AGGCTGTGGG | GCGCAGGGGC | ACCTTCGGGA | GGGAGTGGGT | 9800 |
| 70 | ACCGTGACAG | CCCTGGTCTT | GCAGAGACGC | ACCCAGGTTA | CACACGTGGT | GAGTGCAGGC | GGTGACCTGG | 9870 |
| | CTCCTGTGCT | TCTTTGGAAA | GTCAAGAGTG | GCGGCTCTGT | GGGCCCCAGT | GAGACCCCA | GGAGCTGTGC | 9940 |
| | ACAGGGCCTG | CAGGGCCGAG | GCGGCAGCTG | CCTCCCCAGG | GTGCACCTGA | GCCTGCGGAG | AGCAGGAGCT | 10010 |
| | GCTGAGTGAG | CTGGCCACAC | GCGTTCGCTG | CGGTACAGTT | CCTGCGTGGG | GTTGTTTGGG | ATCGGTGGGA | 10080 |
| | GAATTTGGAT | TTGCTGAGTG | CTGCTGTCTT | GAACACGGA | GATGGCTAGG | AGTGGGTTTC | AGAGTTGATT | 10150 |
| 75 | TTTGTGAATC | AAACTAAAT | CAGGCACAGG | GGACCTGGCC | TCAGCACAGG | GGATTGTCCA | ATGTGGTCTC | 10220 |
| | CCTCAAGGGC | GCCCCACAGA | GCCGCTGGGC | TTGTTTAAAA | GTGCGATTGG | ACGAGGGACG | AGAAACCTTG | 10290 |
| | AAAGCTGTAA | AGGGAACCTT | CAGAAAATGT | GGCCGCCAGG | GGTGGTTTCA | GGTGCTTTGC | TGGGCTGTGT | 10360 |
| | TTGTGAAAAC | CCATTTGGAC | CCGCCCTCCA | AGTCCACCCT | CCAGGTCCAC | CCTCCAGGGC | CGCCCTGGGG | 10430 |

| | | | | | | | | |
|----|-------------|-------------|-------------|------------|-------------|------------|-------------|-------|
| | TGGGGGTATG | CCTGGCGTTC | CTTGTGCCGC | AGCCCGGAGC | ACAGCAGGCT | GTGCACATTT | AAATCCACTA | 10500 |
| | AGATTCACTC | GGGGGGAGCC | CAGGTCCCAA | GCAACTGAGG | GCTCAGGAGT | CCTGAGGCTG | CTGAGGGGAC | 10570 |
| | AGAGCAGACG | GGGAACGCTG | CTTCTGTGTG | GCAAGTTCCT | GAGGGTGCTG | GCCAGGGAGG | TGGCTCAGAG | 10640 |
| 5 | TGTATGTTGG | GGTCCCACCG | GGGGCAGAAC | TCTGTCTCTG | ATGAGTCGGC | AGCCATGTAA | CAGGAAGGGG | 10710 |
| | TGGCCACAGG | GAGCTGGGAA | TGCACCAGG | GAGTCCGCGA | GCTGGCCGAG | GTCCACGGGC | CAGGCCACAG | 10780 |
| | GAAGGGCAGG | GGGACGCCCG | GGGCCACAGC | AGAGGCCGCA | GGGAAGGGAAG | GGGATGCCCA | GGCCAGAGCA | 10850 |
| | GAGGCTACCG | GGCACAGGGG | GGCTCCCTGA | GCTGGGTGAG | CGAGGCTCAT | GACTCGGCGA | GGGAACCTCC | 10920 |
| | TTGACGTGAA | GCTGACGACT | GGTGTGCCCC | AGCTCACAGC | CCAGCCAGGT | CCCGCGCCTG | AGCAGGAACT | 10990 |
| 10 | CAGAACCTTC | CCCTTTGTCT | AAAGCACAGC | AGATGCCCTC | AGGGCATCTA | GGAGAAAACA | GGCAAAAGTCG | 11060 |
| | TTGAGAAACG | TCTTAAAGA | AGGTGGGATG | GTGGCAATTT | CTTGTCAGAG | TTTTAGTCTG | CCCCGGACCA | 11130 |
| | CAGATGAGTC | TATAACGGGA | TTGTGGTGTT | GCCATGGGGA | CACATGAGAT | GGACCATCAC | AGAGGCCACT | 11200 |
| | GGGGCTGCAC | CTCCCATCTG | AGTCCTGGCT | GTCCCGGGTC | CAGGCCAGGT | TCTTGCAATG | TCACCTACCT | 11270 |
| | GTCTGCCCG | GGAGACAGGG | AAAGCACCCC | GAAGTCTGGA | GCAGGGCTGG | GTCCAGGCTC | CTCAGAGCTC | 11340 |
| | CTGCCAGGCC | CAGCACCCCTG | CTCCAATCA | CCACTTCTCT | GGGGTTTTCC | AAAGCATTTA | ACAAGGGTGT | 11410 |
| 15 | CAGGTTACCT | CCTGGGTGAC | GGCCCCGCAT | CTGGGGGCTG | ACATTGCCCC | TCTGCCTTAG | GACCTGGTCT | 11480 |
| | CGAGGTGTCC | CTGAGTATGG | CTGCGTGGTG | AACTTGCGGA | AGACAGTGGT | GAACCTCCCT | GTAGAAGACG | 11550 |
| | AGGCCCTGGG | TGGCACGGCT | TTTGTTCAGA | TGCCGGCCCA | CGGCCATTTC | CCCTGGTGCG | GCCTGTCTGT | 11620 |
| | GGATACCCGG | ACCTTGAGG | TGCAGAGCGA | CTACTCCAGG | TGAGCGCACC | TGGCCGGAAG | TGGAGCCTGT | 11690 |
| 20 | GCCCCGCTGG | GGCAGGTGCT | GCTGCAGGGC | CGTTGCGTCC | ACCTCTGCTT | CCGTGTGGGG | CAGGCCACTG | 11760 |
| | CCAAATCCCAA | AGGGTCAGAG | GCCACAGGGT | GCCCTCTGTC | CCATCTGGGG | CTGAGCAGAA | ATGCATCTTT | 11830 |
| | CTGTGGGAGT | GAGGGTGCTC | ACAACGGGAG | CAGTTTTCTG | TGCTATTTTG | GTAAGAGGAA | ATGGTGCACC | 11900 |
| | AGACCTGGGT | GCAGTGGGT | GTCTTCAGAA | AGCAGTCTGG | ATCCGAACCC | AAGACGCCCG | GGCCCTGCTG | 11970 |
| | GGCGTGAGTC | TCTCAAACCC | GAACACAGGG | GCCCTGCTGG | GCATGAGTCC | CTCTGAACCC | GAGACCCTGG | 12040 |
| 25 | GGCCCTGCTG | GGCGTGAGTC | TCTCCGAACC | CAGAGACTTC | AGGGCCCTTT | TGGGCGTGAG | TCTCTCCGCT | 12110 |
| | GTGAGCCCCA | CACCTCAAGG | CTCATCCACA | GTCTACAGGA | TGCCATGAGT | TCATGATCAC | GTGTGACCCA | 12180 |
| | TCAGGGGACA | GGGCCATGGT | GTGGGGGGGG | TCTCTACAAA | ATTCTGGGGT | CTTGTTTCCC | CAGAGCCCCG | 12250 |
| | GAGCTCAAGG | CCCCGTCTCA | GGCTCAGACA | CAAAATGAAT | GAAGATGGAC | ACAGATGCAG | AAATCTGTGC | 12320 |
| | TGTTTCTTTT | ATGAATAAAA | AGTATCAACA | TTCAGGGCAG | GGCAAGGTGG | CTCACACCTA | TAATCCCGAG | 12390 |
| 30 | ACTTTGGGAG | GCCGAGGTGG | GTGGATCACT | TGAGGCCAGG | AGTTTGAGGC | CAACCTAACC | AACATAGTGA | 12460 |
| | AATTCATTTT | CTACTTAAAA | AATACAAAAA | TTAGCCTGGC | CTGGTGGCAC | ACGCCTGTAG | TCCCGGCTAT | 12530 |
| | GCGGGAGGCT | GAGGCAGGAG | AATCATTTGA | ACCCAGGAGG | CAGAGTTTGC | AGTGAGCCGA | GATCACACCA | 12600 |
| | CTGCACTCCA | GCCTGGGCAA | CAGAGTGAGA | CTTCATCTTA | AAAAAATAAA | AAAAAGTATC | AGCATTCCAA | 12670 |
| | AACCATAGTG | GACAGGTGTT | TTTTTATTCT | GTCTTTCGAT | ATATTTTACT | GGTGTGTGTC | TAGAGGCCGG | 12740 |
| 35 | AACTGGGGGT | GCCTTCTCTT | GAAAGGCACA | CCTTCATGGG | AAGAGAAATA | AGTGGTGAAT | GGTTGTTAAA | 12810 |
| | CCAGCTGGTT | AAACTGGGGT | CCTGTCTGTT | TGAGTTAACA | GTCCAGATCT | GGACTTTGGC | TCTTTCCAGA | 12880 |
| | ATGCTCCCTG | GGGTTTGCTT | CATGGGGGAG | CAGCAGGTGT | GGACACCCCT | GTGATGGGGG | AGCAGCAGGT | 12950 |
| | GCAGACGCCC | TCATGATGGG | GGAGTGGCAG | GTGCAGACAC | CCTTGTGCAT | GGTGGCCAGC | ATGTCCCTGT | 13020 |
| | TGAGGCTCCC | TCCCCACAAG | GATGCCGGTC | TCTGTGCTCT | CCCACAGTCC | CTGCTTCCCT | CTCACAGGCT | 13090 |
| 40 | TACCTGGTCC | TGGCCTCCAC | TGGCTTTGTC | TGCATGATTT | CCACATTTCC | TGGGCTCCCA | GCACCTCTTC | 13160 |
| | GCCTCTCCCA | GGCACCTCTG | CAGTGTCTGG | CATACCAAGT | AGCTGTGAAC | TGTCCACTGC | TTATTTTGCT | 13230 |
| | CCCATGAAA | TGATTTTTTT | AGGACAGGCA | CCCCGTGGTC | CAGCCTCTGG | CACAGCATCA | TGAATGTGTA | 13300 |
| | TTGAAGGACA | AAGGACAGAC | AAACAAATCA | GGAAAATGGG | TTCTCTCTAA | ACACATTGCA | AAGCCACAGA | 13370 |
| | GGGTAGTGCA | GGATGGGTGG | GCATCAGGTC | ATCAGATGTG | GGTCCAATGC | CAGAAATATC | TGTGTCTCCA | 13440 |
| 45 | AAGGCCACTT | GGTCAGAGTG | TGTGCTTGCA | GAGGTGGGTC | TAAAAGCTCA | GCAGTGAGG | CAGTGGTTCG | 13510 |
| | CCATACTCAG | GGTGAATCTA | CATCCTCTGT | GTCTGAAGTA | TACAGCAGAG | GCTTGAAAGG | CATCTGGGAG | 13580 |
| | AAGAAAACAG | GCAAAATGAT | TAAGAAAAGT | GAAAAGGAA | AAGTGGTAAG | ATGGGAATTT | TCTGTCTCAG | 13650 |
| | ATTTTAGTCT | CCCAAACAC | AGCTCAGATG | GTAGAATGTG | GTGAGAACTG | ATGGACAGAA | CAATAGAACA | 13720 |
| | AAACGGGAAG | CTATCTCTC | AGAAACGTGT | GTTAATGTGG | TATGTGGCAC | AGCTGATGGA | AAAGAGAGTG | 13790 |
| 50 | TGTGTGTAAT | TTTTTTTTCT | GAGAAAACCTG | ACTGGAAGCA | AATAAGTTGT | GTCTTTACAG | CATATACCAG | 13860 |
| | AGCAGATTCT | AGGTAGAAGA | GGAGACACAT | GCAAAACAAC | CCAGCAACAG | AAATAAAACA | AAAGACTCAA | 13930 |
| | AGGGAAGGGA | GGTGAACGTT | CCCTGGTTTTG | GTGTTGGGGA | AGGACACACA | GGGAGGCGGA | TGAACCCAGT | 14000 |
| | GAGGCAACGG | GCATTGCTTT | CACTGCAGAG | AAACTCAGCT | TGCCTGAGCC | ACAGTGAAAA | TGGCCATTCC | 14070 |
| | CTGGAGCGTT | TGTGCACGTG | ATTTATTTAA | GGGCCCTGTG | GAGGTCTGTC | ACATTCATCC | TCTCACTTTG | 14140 |
| 55 | TTCTCTAAC | CACCTGAGAG | GTAGAGGAGG | AAAGGCTCCA | GGGGAGCAGC | CGCCCTTGTT | CACCCAGTGT | 14210 |
| | GCAAAGGGCA | TGCATGATTG | CAGCCTGGCC | TCTGTCTCCG | GGGCCCTTGC | TCTGCCCCAG | GACCCACAC | 14280 |
| | AAGTCAGACC | CATAGGCTCA | GGGTGAGCCG | GAGCCCAAGG | TCGTGTTGGG | GATGGCTGTG | AAAGAAGAAA | 14350 |
| | TGGACGTCTG | ATGCACACTT | GGGAAGGTCC | TACCAGCAGC | GTCAAAGAAA | TGCATGTGAA | ACTGACAGCG | 14420 |
| | AGACCCATCC | CTCAAAGAAA | CGCACGTGAA | ACTGATGGCG | AGACCTGTCC | CCATCCCTCA | TGCTGGCTCC | 14490 |
| 60 | TTTTCTGGGC | TTGCCAAGAG | CCAGCATCAG | GTTGAGGCAA | GCTGGAAGAA | CTTTTCTGGA | AAGCAGCTTG | 14560 |
| | TTTGCAATGA | AGTCCTCACA | ATGTCCTGTG | TCTTCCCAGT | AATTCACATT | CTGAAGTGAC | CAGACATTAT | 14630 |
| | CACGGGTCTT | ATTTACCAAT | TCCAGTGTTT | CAGGCAGGGG | GACTTGCCAC | AGCAAGTCAC | GAACCTGCCC | 14700 |
| | AAATACAGGG | CTAAGGAGAT | ATTATGCATC | ACAAAACCTG | CTCTGCCATT | AAACATTTTT | CAAGAATTTT | 14770 |
| | TTGAAGAATG | TTTAATGGCA | CAAAAACGTTT | ATTTCAATGT | AGCAGTGTTC | AAAGCTGGAT | GTAAAGAAAC | 14840 |
| 65 | ACACCCAGGA | AGCCTGCCGT | GAATGTCATG | TGTGTTCTAT | TTTGGACATG | GACATACATG | GGCAGTGAGT | 14910 |
| | GGTGGTGAGG | CCCTGGAGGA | CATCGGTGGG | ATGCCCTCAT | CCTGCCCTCT | TGGAGACACC | ATGTGTGCCA | 14980 |
| | CGTGCACTCA | CTGGAGCCCT | GTTTAGCTGG | TGCCACCTGG | CTCTTCCATC | CCTGAGATTC | AAACACAGTG | 15050 |
| | AGATTCCCCA | CGCCCAACTC | AGTGTCTCTC | CACAAAAAAC | CTGAGTCACA | CCTGTGTTCA | CTCGAGGGAC | 15120 |
| | GCCCGGGAGC | CAGGGCTCCA | CAGTTTATTA | TGTGTTTTTG | GCTGAGTTAT | GTGCAGATCT | CATCAGGGCA | 15190 |
| 70 | GATGATGAGT | GCACAAACAC | GGCCGTGCGA | GGTTGGGATA | CACTCAACAT | CACTAGCCAG | GTCTGGTGG | 15260 |
| | AGTTTGGTCA | TGCAGAGTCT | GGATGGCATG | TAGCATTTGG | AGTCCATGGA | GTGAGCACCC | AGCCCCCTCG | 15330 |
| | GGGTGACGCG | CATGCCCCAG | GCAGGACAAG | GAAGCGGGAG | GGCTCTTTGG | GGCTCTTTGG | AGCAAGCTTT | 15400 |
| | GCAGGAGGGG | GCTGGGTGTG | GGGCAGGCAC | CTGTGCTGTA | CATTCCCCCC | TGTGTCTCAG | CTATGCCCGG | 15470 |
| | ACCTCCATCA | GAGCCAGTCT | CACCTTCAAC | CGCGCTTCA | AGGCTGGGAG | GAACATGCGT | CGCAAACTCT | 15540 |
| 75 | TTGGGGTCTT | CGCGGTGAAG | TGTCAAGACC | TTTCTGGA | TTTGCAGGTG | AGCAGGCTGA | TGGTCAAGCT | 15610 |
| | AGAGTTCAGA | GTTCAAGAGG | TGTGTGCGCA | AGTATGTGTG | TGTGTGTGTG | CGCGCGTGCC | TGCAAGGCTG | 15680 |
| | ATGGTGACTG | GCTGCACGTA | AGAGTGCACA | TGTACGCATA | TACACGTGAG | CACATACATG | TGTGCATGTG | 15750 |
| | TGTACATGAA | GGCATGGCAG | TGTGTGCACA | GGTGTGCAAG | GGCACAAGTG | TGTGCACATG | CGAATGCACA | 15820 |

| | | | | | | | | |
|----|------------|-------------|------------|------------|------------|------------|-------------|-------|
| | CCTGACATGC | ATGTGTGTTC | GTGCACAGTC | GTGTGGGCAT | TCACGTGAGG | TGCATGCGTG | TGGGTGTGCA | 15890 |
| | GTGTGAGTAG | CATGTGTGCA | CATAACATGT | ATTGAGGGGT | CCTCGTGTTC | ACCCCGCTAG | GTCTCTCAGCA | 15960 |
| | CCAGTGCCAC | TCCTTACAGG | ATGAGACGGG | GTCCAGGGC | TGGTGGGCT | GAGGCTCTGA | AGCTGCAGCC | 16030 |
| 5 | CTGAGGGCAT | TGTCCCATCT | GGGCATCCGC | GTCCACTCCC | TCTCCTGTGG | GCTTCTGTGT | CCACTCCCC | 16100 |
| | TCTCCTGTGG | GCATTTACAT | CCACTCCACT | CCCTCTCTCC | TGTGGGCATC | CGCGTCCACT | CCCCCTCTCT | 16170 |
| | GTGGCATCTG | GGGTCCACCT | CCCCTCTCTG | TGGGCATTTG | CGTCCACTCC | CTCTCCTGGT | TCCTCTCTGT | 16240 |
| | CTTGGCCGAG | CCTCGGGGGC | AGGCAGATGA | CACAGAGTCT | TGACTCGCCC | AGGGTGGTTC | GCAGCTGCCG | 16310 |
| | GGTGAGGGCC | AGGCCCGATT | TCACTGGGAA | GAGGGATAGT | TTCTTGTCAA | AATGTTCTCT | TTTCTTGTTC | 16380 |
| 10 | CATCTGAATG | GATGATAAAG | CAAAAAGTAA | AAACTTAAAA | TCCCAGAGAG | GTTTCTACCG | TTTCTCACTC | 16450 |
| | TTTCTTGGCG | ACTCTAGGTG | AACAGCCTCC | AGACGGTGTG | CACCAACATC | TACAAGATCC | TCCTGTCTGA | 16520 |
| | GGCGTACAGG | TGAGCCGCCA | CCAAGGGGTG | CAGGCCACGC | CTCCAGGGAC | CCTCCGCGCT | CTGCTCACCT | 16590 |
| | CTGACCCGGG | GCTTCACCTT | GGAACCTCTG | GGTTTTAGGG | GCAAGGAATG | TCTTACGTTT | TCAGTGGTGC | 16660 |
| | TGCTGCCCTG | GCACAGTTCT | GTTCCGCTGG | CTCTGTGCAA | AGCACCTGTT | CTCCATCTCT | GGGTAGTGGT | 16730 |
| | AGGAGCCGGT | GTGGCCCGAG | GTGTCCCCAC | TGTGCCTGTG | CACTGGCCGT | GGGACGTCAT | GGAGGCCATC | 16800 |
| 15 | CCAGGGCAGC | AGGGGCATGG | GGTAAAGAGA | TGTTTATGGG | GAGTCTTAGC | AGAGGAGGCT | GGGAAGGTGT | 16870 |
| | CTGAACAGTA | GATGGGAGAT | CAGATGCCCG | GAGGATTTGG | GGTCTCAGCA | AAGAGGGCCG | AGGTGGGTGC | 16940 |
| | AGGTGAGGGT | CGCTGGCCCC | ACCCCGGGGA | AGGTGCAGCA | GAGCTGTGGC | TCCCCACACA | GCCCGGCCAG | 17010 |
| | CACCTGTGCT | CTGGGCATGG | CTGTGCTCCT | GGAACGTTCC | CTGTCTGTGG | TGGTCAGGGG | GTGCCCTGTC | 17080 |
| | CAAGAATCGA | CAACTTTATC | ACAGAGGGAA | GGCCCAATCT | GTGGAGGCCA | CAGGGCCAGC | TTCTGCTGGT | 17150 |
| 20 | AGTCAGGGCA | GGTGGTGCCA | CAAGCCTCGG | GGCTGTACCA | AAGGGCAGTC | GGGCACCACA | GGCCCGGGCC | 17220 |
| | TCCACCTCAA | CAGGCCCTCC | GAGCCACTGG | GAGCTGAATG | CCAGGAGGCC | GAAGCCCTCG | CCCCATGAGG | 17290 |
| | GCTGAGAAGG | AGTGTAGACA | TTTGTGTTAC | CCAGGGCCGA | GGCTGCGCGA | ATTACCGTGC | ACACTTGATG | 17360 |
| | TGAAATGAGG | TCGTGCTCTA | TCGTGGAAGC | CCAGCAAGGG | CTCACGGGAG | AGTTTTCAT | TACAAGGTCT | 17430 |
| | TACCATGAAA | ATGGTTTTTA | ACCCGAGTGC | TTGCGCCTTC | ATGCTCTGGC | AGGGAGGGCA | GAGCCACAGC | 17500 |
| 25 | TGCATGTTAC | CGCCTTTGCA | CCAGCTCCAG | AGGCTTGGGA | CCAGGCTGTC | TCAGTTCCAG | GSTGCGTCCG | 17570 |
| | GCTCAGACCG | CCCTCCTCTC | TGCCCTTCTC | CTCTGCCTCA | AATCTTCCCT | CGTTTGCATC | TCCTGTGACG | 17640 |
| | GTGCCTGGGC | CCTCGTGCAA | GCTGCTTGAC | TCCTTTCCGG | AAACCTTTGG | GGTGTGCTGG | ATACAGGTGC | 17710 |
| | CACTGAGGAC | TGGAGGTGTC | TGACACTGTG | GTGACCCCA | GGGTCCAGCT | GGCGTGCTTG | GGGCCTCCTT | 17780 |
| 30 | GGCGGCTGAT | GAGGTACAGG | GAGTTTTCCC | AGGTGAAAC | TCCTGGGAAA | CTCCAGGGC | CATGTGACCT | 17850 |
| | GCCACCTGCT | CCTCCCATAT | TCAGCTCAGT | CTTGTCTCTA | TTTCCCCACC | AGGGTCTCTA | GCTCCGAGGA | 17920 |
| | GCTCCCTGAT | AGGGCCTGGG | CTCAGGGCAG | GGCGGCTGAG | TTTCCCCACC | CATGTGGGGA | CCCTTGGGTA | 17990 |
| | GTGCTTGAT | TGGGTAGCCC | TGAGGAGGCC | GAGATGCCAT | GGGCCACGGG | CCGTTTCCAA | ACACAGAGTC | 18060 |
| | AGGCACGTGG | AAGGCCCAGG | AATCCCTTTC | CCTCGAGGCA | GGAGTGGGAG | AACGGAGAGC | TGGGCCCGCA | 18130 |
| 35 | TTTCACGGCA | GCCAGGCTGC | AGTGGGCGAG | GCTGTGGTGG | TCCACGTGGC | GCTGGGGGGC | GGGTCTGATT | 18200 |
| | CAAACTCCGT | GGGGCTCGGC | CTTCTCTGGC | CGTGTGGGCA | GCGCCTCCAC | ACGGGCTTGG | GGTGGAGGCT | 18270 |
| | CCGACCTCTA | CGAGGTGGCT | ATTCTCCCTT | TTGGAAGAGA | GCCCCCACC | CATGCTAGGT | TTTCCCTGCC | 18340 |
| | TGGGTACAGG | GCGTGGCCGT | GTGGCAACCC | CGGGACCTTA | GGCTTATTTA | TTTGTTTAAA | AACATTCTGG | 18410 |
| | GCCTGGCTTC | CGTTGTTGCT | AAATGGGGAA | AAGACATCCC | ACCTCAGCAG | AGTTACTGAG | AGGCTGAAAC | 18480 |
| 40 | CGGGGTGCTG | GCTTGACTGG | TGTGATCTGA | GGTCATTCCA | GAAGTGGCTC | AGGAAGTCAG | TGAGACAGG | 18550 |
| | TACATGGGGG | GCTCAGGCAG | TGGGTGAGAT | GAGGTACACG | GGGGGCTCAG | GCAGTGGGTG | AGGCCAGGTA | 18620 |
| | CATGGGGGGC | TCAGGCACCTG | GGTGAGATGA | GGTACACGGG | GGGCTCAGGC | AGAGGGTCAG | ACCAGGTACA | 18690 |
| | CGGGGGCTCT | GACATACAGC | ACATATGAGC | ACATGTGCAC | ATGTGCTGTT | TCATGGTAGC | CAGGTGCTGT | 18760 |
| | CACACCTGCC | CCAAAGTCCC | AGGAAGCTGA | GAGGCCAAAG | ATGGAGGCTG | ACAGGGCTGG | CGCGGTGGCT | 18830 |
| 45 | CACACCTGTA | GTCCACAGCA | TTTGGGAGGC | CGAGGCGAGA | GGATCCCTTG | AGCCCAGGAG | TTTAAGACCA | 18900 |
| | GCCTGAGCAA | CATAGTAGAA | CCCCATCTCT | ATGAAAAATA | AAAACAAAAA | TTAGCTGAAC | ATGTTGGTGT | 18970 |
| | GCGCCTGTAG | TTCCAATACT | TGGGAGGCTG | AAGTGGGAGG | ATCACTTGAG | CCCAGGAGGT | GGAAGCTGCA | 19040 |
| | GTGAGCTGAG | ATTGCACCA | TGTACTGCAG | CCTGGGTGAG | AGAGTGAAG | CCCATCTCAA | CAACAACAAA | 19110 |
| | GAAGACTGAC | AAATGCAATT | TCTTGGAAAG | AAACATTTC | TAGGAACCTA | ACCTACACAC | AGAAGCTAAG | 19180 |
| 50 | TCGGTGTCTC | GGTGTCAAGT | AGATGAGATG | ATGGGCTCTC | ACACCATCAC | CCAGACCCA | AGGGTTATGC | 19250 |
| | ACCACAGGGG | CGGGTGGCTC | AGAAGGGATG | CGCAGGACGT | TGATATACGA | TGACATCAAG | GTTGTCTGAC | 19320 |
| | GAAGGGCAGG | ATTGATGATA | AGTACCTGCT | GGTACACAA | GAACAATGGA | TAAACTGGAA | ACCTTAGAGG | 19390 |
| | CCTTCCCGGA | ACAGGGGCTA | ATCAGAAGCC | AGCATGGGGG | GCTGGCATCC | AGGATGGAGC | TGCTGACGCC | 19460 |
| | TCCACATGCG | TGTTTCATACA | GATGGTGCAC | AGAAACGCAG | TGTACCTGTG | CACACACAGA | CACGCAGCTA | 19530 |
| 55 | CTCGCACACA | CAAGCACACA | CACAGACATG | CATGCATGCA | TCCGTGTGTG | TGCACCTGTG | CCCATGAGGA | 19600 |
| | AACCCATGCA | TGTGCATTCA | TGCACGCACA | CAGGCACCCG | TGGGCCCATG | CCCACACCCA | CAGCACCCGT | 19670 |
| | CTGATTAGGA | GGCCTTTCTC | CTGACGCTGT | CCGCCATCCT | CTCAGGTTTC | ACGCATGTGT | GCTGCAGCTC | 19740 |
| | CCATTTTCAT | AGCAAGTTTG | GAAAGACCCC | ACATTTTTC | TGCGCGTCAT | CTCTGACACG | CGCTCCCTCT | 19810 |
| | GCTATCCAT | CCTGAAAGCC | AAGAACGCAG | GTATGTGCAG | GTGCTGGCC | TCAGTGGCAG | CAGTGCCTGC | 19880 |
| 60 | CTGCTGGTGT | TAGTGTGTCA | GGAGACTGAG | TGAATCTGGG | CTTAGGAAGT | TCTTACCCCT | TTTCGCATCA | 19950 |
| | GGAAGTGGTT | TAACCCAACC | ACTGTACGGC | TCGTCTGCC | GCCCTCTCGT | GGGGTGAACA | GAGCACCTGA | 20020 |
| | TGGAAGGGAG | AGGAGCTGTC | TGGGAGCTGC | CATCCTTCCC | ACCTTGCTCT | GCCTGGGGAA | GCGCTGGGGG | 20090 |
| | GCCTGGTCTC | TCTGTTTGG | CCCATGGTGG | GGTTTGGGGG | GCTTGGCCTC | TCCTGTTTGC | CCTGTGGTGG | 20160 |
| | GATTGGGCTG | TCTCCCGTCC | ATGGCACTTA | GGGCCCTTGT | GCAAACCCAG | GCCAAGGGCT | TAGGAGGAGG | 20230 |
| 65 | CCAGGCCAGG | GCTACCCAC | CCCTCTCAGG | AGCAGAGGCC | GCGTATCACC | ACGACAGAGC | CCCGCGCCGT | 20300 |
| | CCTCTGCTTC | CAGTCAACCG | TCCTCTGCC | CTGGACACT | TGTCCAGCAT | CAGGGAGGTT | TCTGATCCGT | 20370 |
| | CTGAAATTCA | AGCCATGTCC | AACCTGCGGT | CCTGAGCTTA | ACAGCTTCTA | CTTTCTGTTC | TTTCTGTGTT | 20440 |
| | GTGGAAATTT | CACCTGGAGA | AGCCGAAGAA | AAATTTTCTG | TCGTGACTCC | TGCGGTGCTT | GGGTGGGGAC | 20510 |
| | AGCCAGAGAT | GGAGCCACCC | CGCAGACCGT | CGGGTGTGGG | CAGCTTTCCG | GTGTCTCCTG | GGAGGGGAGC | 20580 |
| 70 | TGGGCTGGGC | CTGTGACTCC | TCAGCCTCTG | TTTTCCCCCA | GGGATGTGCG | TGGGGGCCAA | GGGCGCCGCC | 20650 |
| | GGCCCTCTGC | CCTCCGAGGC | CGTGCAGTGG | CTGTGCCACC | AAGCATTCCT | GCTCAAGCTG | ACTCGACACC | 20720 |
| | GTGTCACTTA | CGTGCCACTC | CTGGGGTCC | TCAGGACAGG | CAAGTGTGGG | TGGAGGCCAG | TGCGGGCCCC | 20790 |
| | ACCTGCCAGG | GGGTATCCTT | TGAACGCCCT | GTGTGGGGCC | AGCAGCCTCA | GATGCTGCTG | AAGTGCAGAC | 20860 |
| | GCCCCCGGGC | CTGACCCCTG | GGGCCCTGGG | CCACGCTGGC | AGCCCTATGT | GATTAACAGC | TGGTGTCCCC | 20930 |
| | AGGCCACGGA | GCCTGGCAGG | GTCCCCAACT | TCTTGAACCC | CTGCTTCCCA | TCTCAGGGGC | GATGGCTCCC | 21000 |
| 75 | CACGCTTGGG | AGCCTTCTGA | CCCCTGACCT | GTGCTCTCTC | ACAGCCTCTT | CCCTGGCTGC | TGCCCTGAGC | 21070 |
| | TCCTGGGGTC | CTGAGCAAGT | TCTCTCCCCG | CCCCGCCGCT | CCAGCGTCAC | TGGGCTGCCT | GTCTGCTCGC | 21140 |
| | CCCGGTGGAG | GGGTGCTGCT | CCCTTCACTG | AGTTTCCAC | CAGCCAGGGC | CACGAGGTGC | AGGCCCTGCC | 21210 |

| | | | | | | | | | | | |
|------------|------------|-------------|-------------|--------------|------------|------------|-------|----|----|----|----|
| | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 |
| TCGCCCGCCA | CCACACAGCT | CTAGGAGGCT | TGGAGGATGC | CACCTCTGGC | CTCTTCTGGA | ACGGAGTCTG | 21280 | | | | |
| ATTTTGGCCC | CGCAGCCGAC | ACGCAGCTGA | GTCGGAGGCT | CCCGGGGACG | ACGCTGACTG | CCCTGGAGGC | 21350 | | | | |
| CGCAGCCAAC | CCGGCAGCTG | CCTCAGACTT | CAAGACCATT | CTGGACTGAT | GGCCACCCCG | CCACAGCCAG | 21420 | | | | |
| GCCGAGAGCA | GACACACGCA | GCCTGTACAC | CGCGGGCTCT | ACGTCCCGAT | GAGGAGGGGG | CGGCCACAC | 21490 | | | | |
| CCAGGCCCGC | ACCGCTGGGA | GTCTGAGGCC | TGAGTGAGTG | TTTGGCCGAG | GCCTGCATGT | CCGGCTGAAG | 21560 | | | | |
| GCTGAGTGTC | CGGCTGAGGC | CTGAGCGAGT | GTCCAGCCAA | GGGCTGAGTG | TCCAGCACAC | CTCCCGTCTT | 21630 | | | | |
| CAGTTCCCCA | CAGGCTGGGC | CTCGGCTCCA | CGGGGGGCC | AGGTTTTCCT | CACCAAGAGC | CCGGCTTCCA | 21700 | | | | |
| CTCCCCACAT | AGGAATAGTC | CATCCCCAGA | TTCGCCATTG | TTCACCCCTC | GCCCTGCCCT | CCTTTGCCCT | 21770 | | | | |
| CCACCCCCAC | CATCCAGGTG | GAGACCCTGA | GAAGGACCCT | GGGAGCTCTG | GGAAATTTGA | GTGACCAAAG | 21840 | | | | |
| GTGTGCCCTG | TACACAGGCG | AGGACCCCTG | ACCTGGATGG | GGGTCCCTGT | GGGTCAAATT | GGGGGAGAGT | 21910 | | | | |
| GCTGTGGGAG | TAAAATACTG | AATATATGAG | TTTTTCAGTT | TTGAAAAAAA | TCTCATGTTT | GAATCCTAAT | 21980 | | | | |
| GTGCATGCA | TAGACACCAC | TGTATGCAAT | TACAGAAGCT | TGTGAGTGAA | CGGGTGGTGT | GTGAGTCGGG | 22050 | | | | |
| GCCCATGGCC | TGGCTGTGCA | TTTACGGGAG | TCTATGAGTG | AATGGGTTTG | TGGTCACTGC | GGGCCCATAG | 22120 | | | | |
| CCTGGCTGGG | CCTGGGAGGT | TTCTGATGCT | GTGAGGCAGG | AGGGGAAGGA | GGGTAGGGGA | TAGACAGTGG | 22190 | | | | |
| GAGCCCCCAC | CCTGGAAGAC | ATAACAGTAA | GTCCAGCGCC | GAGGGGACGC | AGGAGTCTGT | GGGGCCCCAG | 22260 | | | | |
| TTGGGCGGCG | GGGATGATGG | AGGGCTGGCA | CAGGGTGGCA | GGGATGATGC | GGGCCCCAGC | TGGGTTGGCA | 22330 | | | | |
| GGGGTGATGG | GGGGGGCTGG | TCTGGGTGGC | GGGGAAGATG | GGGAAGCCTG | GCTGGGCCCC | CTCTCTCCCT | 22400 | | | | |
| GCCTCCCCAC | TGCAGCCGCT | GATCCGGATG | TGCTTCCCTG | GTGCACATCC | TCTGGGCCAT | CAGCTTTTCT | 22470 | | | | |
| GGAGTGGGGG | GGCAGGGGCA | TGACACCATC | CTGTATAAAA | TCCAGGATTC | CTCTCTCTGA | ACGCCCCAAC | 22540 | | | | |
| TCAGGTTGAA | AGTCACATTC | CGCCTCTGGC | CATTCTCTTA | AGAGTAGACC | AGGATCTCTG | TCTCTGAAGG | 22610 | | | | |
| GTGGGTFAGG | TGGGGCAGTG | GAGGGTGTGG | ACACAGGAGG | CTTCAGGGTG | GGGCTGGTGA | TGCTCTCTCA | 22680 | | | | |
| TCTCTTATC | ATCTCCCACT | CTCATCTCTC | ATCCTCTTAT | CATCTCCGAC | TCTCATCTGT | CTTCTCTTAT | 22750 | | | | |
| TCTCCAGTC | TCATCTGTCA | TCCTCTTACC | ATCTCCCACT | CTCATCTCTT | ATCTCTTATC | CTCTAGTCTT | 22820 | | | | |
| CATCCAGACT | TACCTCCCCA | GGCGGGTGCC | AGGCTCGCAG | TGGAGCTGGA | CATACGTCTT | TCCTCAGGCA | 22890 | | | | |
| GAAGGAACCT | GAAAGGATGC | AGAGAACAGG | AGGGGCGGCT | CAGAGGGACG | CAGTCTTTGG | GTGAAGAAAC | 22960 | | | | |
| AGCCCCCTCT | CAGAAGTTGG | CTTTGGGCCAC | ACGAACCCGA | GGGCCCTCGG | TGAGTGGCTG | CAGAGCCTTC | 23030 | | | | |
| CAGCAGGTCC | CTGGTGGGGC | CTTATGGTAT | GGCCGGGTCC | TACTGAGTGC | ACCTTGAGCA | GGGCTTCTGG | 23100 | | | | |
| TTTGAGTGCA | CCCGGACAGT | GCCTGGTGTC | GGGGTGGGGG | CTTATGGCCA | CTGGATATGG | CGTCATTTAT | 23170 | | | | |
| TGCTGTGCT | TCCAGAGAAT | TCFGAGTGAC | GAGCCTTAAT | GTGTATGGTG | GGCCCAAGTC | CACAGACTGT | 23240 | | | | |
| GTGCTAAATG | CACCTCTGGT | CTTGGAGGCC | CCGTATAGGA | GCTGTGAGGA | AGGAGGGGCT | CTTGGCAGCC | 23310 | | | | |
| GGCCTGGGGG | CGCCTTTGGC | CTGCAAACTG | GAAGGGAGCG | GCCCCGGGCG | CCGTGGGCGG | ACGACCTCAA | 23380 | | | | |
| GTGAGAGGTT | GGACAGAACA | GGCGGGGGAC | TCCCAAGGAG | CAGAGGCGCG | TGCTCAGGCA | CACCTGGGTT | 23450 | | | | |
| TGAATCAGAG | ACCAACaGTT | CAGGCCATTG | TTCAGCTATC | CATCTTCTAC | AAAGCTCCAG | ATTCTCTGTT | 23520 | | | | |
| CTCCGGGTGT | TTTTTGTGTA | AATTTTACTC | AGGATTACTT | ATATTTTATT | CTAAAGTATT | AGACCTTTAA | 23590 | | | | |
| AAAAGGTATT | TGCTTTGATA | TGGCTTAACT | CACATAAGCA | CTACTTTTAT | TGCTGTGTTT | TATTTATTAT | 23660 | | | | |
| TATTATTATT | ATTAGAGATC | GTGTCTACTC | TGTCACCCAG | GTGTGTTAGT | CAGTGGCACA | GTGATGGCTC | 23730 | | | | |
| GCTGTAGCCG | CAAAACCCCA | GGCTCAAGTG | ATCCTCCGGC | CTCAGCTTCC | CAGAGTCTGT | GGATTACAGG | 23800 | | | | |
| TGTGAGCCAG | TGCCCTTGCC | TGGCACTTTT | AAAAAACCACT | ATGTAAGTCT | AGGTCAGCTG | GCTTCCACAC | 23870 | | | | |
| CTGTGATCCC | AGTAGTTTGG | GAAGCCGAGG | CAGAAGGATT | GTCTGAGGCC | AGGAGTTTGA | GACCAGCATG | 23940 | | | | |
| GGTAACATAG | GGAGACCCCA | TCTCTACAAA | AAATGCAAAA | AGTTATCCGG</ | | | | | | | |

60 **Beispiel 5**

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3' untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in

5 den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

| | 5'-Exon | | | Intron | | | | 3'-Exon | | | |
|----------------|---------|----|----|--------|-----|-----|-----|---------|-----|-----|------------|
| Prä-mRNA | A/C | A | G | | G | T | A/G | A | ... | N C | A G G |
| 10 Häufigk.(%) | 70 | 60 | 80 | | 100 | 100 | 95 | 70 | | 80 | 100 100 60 |

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616

15 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

Intron 1: bp 11493-11596 (SEQ ID NO 4);

Intron 2: bp 12951-21566 (SEQ ID NO 5);

Intron 3: bp 21763-23851 (SEQ ID NO 6);

5 Intron 4: bp 24033-24719 (SEQ ID NO 7);

Intron 5: bp 24900-25393 (SEQ ID NO 8);

5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

10 Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden Positionen lokalisiert:

3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);

Intron 7: bp 3879-4858 (SEQ ID NO 11);

Intron 8: bp 4945-7429 (SEQ ID NO 12);

Intron 9: bp 7544-9527 (SEQ ID NO 13);

15 Intron 10: bp 9600-11470 (SEQ ID NO 14);

Intron 11: bp 11660-15460 (SEQ ID NO 15);

Intron 12: bp 15588-16467 (SEQ ID NO 16);

Intron 13: bp 16530-19715 (SEQ ID NO 17);

Intron 14: 19841-20621 (SEQ ID NO 18);

20 Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position 21960-25138 (SEQ ID NO 20).

25 Die genannten Introns haben im einzelnen folgende Sequenzen:

Intron 1 (SEQ ID NO 4)

GTGGGCCTCCCCGGGTCGGCGTCCGGCTGGGGTTGAGGGCGGCCGGGGGAACCAGCGACATGCGGAGAGCAGCGCAGG
CGACTCAGGGCGCTTCCCCCGCAG

5

Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCCGTGAGGGCCAGGCCCCAGAGCTGAATGCAGTAGGGGCTCAGAAAAGGGGGCAGGCAGAGCC
CTGGTCCTCCTGTCTCCATCGTCACGTGGGCACACGTGGCTTTTCGCTCAGGACGTGAGTGACACGGTGATCTCTGCC
TCTGCTCTCCCTCCTGTCCAGTTTGATATAAATTACGAGGTTACCTTCACGTTTTGATGGACACGCGGTTTCCAGGCGC
10 CGAGGCCAGAGCAGTGAACAGAGGAGGCTGGGCGCGGCAGTGGAGCCGGTTGCCGGCAATGGGGAGAAGTGTCTGGAAG
CACAGACGCTCTGGCGAGGGTGCCTGCAGGTTACCTATAATCCTCTTCGCAATTTCAAGGGTGGGAATGAGAGGTGGGGA
CGAGAACCCCTCTTCTGGGGTGGGAGGTAAGGGTTTTGCAGGTGCACGTGGTCAGCCAATATGCAGGTTTGTGTTTA
AGATTTAATGTGTGTTGACGGCCAGGTGCGGTGGCTCACGCCGTAATCCCAGCACTTTGGGAAGCTGAGGCAAGTGA
TCACCTGAGGTGAGGAGTTGAGACCAGCCTGACCAACATGGTGAAACCTATCTGTACTAAAAATACAAAAATTAGCTG
15 GGCATGGTGGTGTGTGCTGTAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGAGGC
TGCAGTGAGCTGAGATTGTGCCATTGTACTCCAGCTGGGCGACAAGAGTGAACTCTGTCTTTAAAAAAGTGT
CGTTGATTGTGCCAGGACAGGGTAGAGGGAGGAGATAAGACTGTTCTCCAGCACAGATCCTGGTCCCATCTTTAGGTAT
GAAGAGGGCCACATGGGAGCAGAGGACAGCAGATGGCTCCACCTGCTGAGGAAGGGACAGTGTGTTGTGGTGTTCAGGGG
ATGGTGCTGCTGGGCCCTGCCGTGTCCCCACCCTGTTTTCTGGATTGATGTTGAGGAACCTCCGCTCCAGCCCCCTT
20 TGGCTCCAGTGCTCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCCGATTTACCCCCCTCCCAAACTCCCAAGAC
ATGTAAGACTTCCGGCCATGCAGACAAGGAGGGTGACCTTCTTGGGGCTCTTTTTTTCTTTTTTTCTTTTATGGTGGC
AAAAGTCATATAACATGAGATTGGCACTCCTAACACCGTTTTCTGTGTACAGTGAGAATTGCTAACTCGGCGGTGTTTA
CAGCAGGTTGCTTGAAATGCTGCGTCTTGCCTGACTGGAAGTCCCTACCCATCGAACGGCAGCTGCCTCACACCTGCTGC
GGCTCAGGTGGACCACGCCGAGTCAGATAAGCGTCATGCAACCCAGTTTTGCTTTTTGTGCTCCAGCTTCCTTCGTTGAG
25 GAGAGTTTGAGTTCTCTGATCAGGACTCTGCCTGTCTATTGCTGTTCTCTGACTTCAGATGAGGTGACAATCTGCCCCCTGG
CTTATGCAGGGAGTGAGGCGTGGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGTTGCCCCAGGTGTCCCT
GTCACGTGTAGGGTGAGTGAGGCGCGCCCCGGGTGTCCCTGTCCGTGCAGCGTGATTGAGGTGTGGCCCCGGGTGT
CCCTGTACGTGTAGGGTGAGTGAGGCGCCATCCCCGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCGGTGTCCCCG
GTGTCCCTGTCCGTGCAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGCGGTCC
30 CCGGGTGTCCCTCTCAGGTGTAGGGTGAGTGAGGCGCGCCCCAGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCACC
GTCCCTGGGTGTCCCTCCAGGTATAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGG
CGCGGCCCGGGTGTCCCTCTCAGGTGCAGGGTGAGTGAGGCGCTGTCCCTGGGTGTCCCTGTCTCGTGTAGGGTGAGT
GAGGCTCTGTCCCCAGGTGTCTTGGCGTTTGCTCACTTGAGCTTGCTCCTGAATGTTTGCTCTTTCTATAGCCACAGCT
GCGCCGGTTGCCCATTGCTGGGTAGATGGTGACGGCGCAGTGCTGGTCCCCAAGCCTATCTTTCTGATGCTCGGCTCT
35 TCTTGGTCACCTCTCCGTTCCATTTTGCTACGGGGACACGGGACTGCAGGCTCTCGCTCCCGGTGCCAGGCACTGCAG
CCACAGCTTCAGGTCCGCTTGCTCTGTTGGGCCTGGCTTGCTCACCACGTGCCCGCCACATGCATGCTGCCAATACTCC
TCTCCAGCTTGCTCTCATGCCGAGGCTGGACTCTGGGCTGCCTGTGTCTGCTGCCACGTGTTGCTGGAGACATCCCAGAA
AGGGTTCTCTGTGCCCTGAAGGAAAGCAAGTCACCCCAGCCCCCTCACTTGCTCTGTTTTCTCCCAAGCTGCCCTCTGC
TTGGCCCCCTTGGGTGGGTGGCAACGCTTGCTCACCTTATCTGGGCACCTGCCGCTCATTGCTTAGGCTGGGCTCTGCCT
40 CCAGTCGCCCCCTCACATGGATTGACGTCCAGCCACAGGTGGAGTGTCTGTCTGTCTCTGCTCTGAGACCCACGTG

5
10
15
20
25
30

GGCCCTCGGTGAGCTGGAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCTGTGAGCTGGATG
TGTGGTGTCTGGATGGTGCAGGTGAGGGTGAGGTCTCCAGGCCCTCGGTAAGCTGGAGGTATGGAGTCCGGATGATGCA
GGTCCGGGGTGAGGTCCGACAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCTGGGGTGAGGTACCC
AGGCCCTGCGGTGAGCTGGGTGTGCGGTGTCTGGATGGTGCAGGTCTGGAGTGAGGTCCGACAGCGGTGCCAGACCATGC
GGTGAGCTGGATATGCGGTGTCCGGATGGTGCAGGTCTGGGGTGAGGTGCCAGGCCCTGCTGTGAGTTGGATGTGGGGT
GTCCGGATGCTGCAGGTCCGGTGTGAGGTACCCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCT
GGGGTGAAGGTCCGACAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCTGGAGTGAGGTCCGACAG
GCCCTCGGTGAGCTGGATGTGCAGTGTCCAGATGGTGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCGGTGAGCTGGATG
TGCGGTGTCTGGATGGTGCAGGTCTGGAGTGAGGTCCGACAGGCCCTCGGTGAGCTGGATGTATGGAGTCCGGATGGTGCC
GGTCCGGGGTGAGGTCCGACAGGCCCTGCTGTGAGCTGGATGTGCGGTGTCTGGATGGTACAGGTCTGGAGTGAGGTCCGAC
AGGCCCTGCTGTGAGCTGGATATGCGGTGTCCGGATGGTGCAGGTGAGGGTGAGGTCTCCAGGCCCTCGGTGAGCTGGA
GGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCTGTGAAGTGGATGTGCGGCGTCTGGATGGT
GCAGGTCTGGGGTGTGGTCCGACAGGCCCTCGGTGAGCTGGAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCCG
CCAGGCCCTGCTGTGAGCTGGATGTGCGGCGTCTGGATGGTGCAGGTCTGGGGTGTGGTCCGACAGGCCCTCGGTGAGCTG
GAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATG
GTGCAGTCCGGGGTGAGGTCCGACAGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCTGGGGTGAGGT
CACCAGGCCCTGCGGTGAGCTGGTGTGCGGTGTCCGGTGTCTGCAGGTCCGGGGTGAGTTCGCCAGGCCCTCGGTGAGC
TGGATGTGCGGTGTCCCGTGTCCGGATGGTGCAGGTCCAGGGTGAGGTCCGCTAGGCCCTTGGTGGGCTGGATGTGCCGT
GTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGACAGGCCCTTGGTGAGCTGGATGTGCGGTGTCTGCATGGTGCAGGTCTG
GGGTGAGGTCCGACAGGCCCTTGGTGGGCTGGATGTGTGGTGTCCGGATGGTGCAGGTCCGGCGTGAGGTCCGACAGGCCCT
GCTGTGAGCTGGATGTGCGGTGTCTGGATGGTGCAGGTCCGGGGTGAGGTAGCCAAGGCCCTCGGTGAGCTGGATGTGGG
GTGTCCGGATGGTGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCGGTTAGCTGGATATGCGGTGTCCGGATGGTGCAGGT
CCGGGGTGAGGTACCCAGGCCCTGCGGTTAGCTGGATGTGCGGTGTCTGGATGGTGCAGGTCCGGGGTGAGGTCCGACAGG
CCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCAGTGAGCTGGATG
TGCTGTATCCGGATGGTGCAGGTCTGGCGTGAGGTCCGACAGGCCCTGCGGTTAGCTGGATATGCGGTGTCCGGATGGTGC
GGTCCGGGGTGAGGTACCCAGGCCCTGCGGTTAGCTGGATGTGCGGTGTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGAC
AGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCGGTGAGCTGG
ATGTGCTGTATCCGGATGGTGCAGGTCTGGCGTGAGGTCCGACAGGCCCTGCGGTGAGCTGGATGTGCAGTGTACGGATGG
TGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCGGTGGGCTGTATGTGTGTGTCTGGATGGTGCAGGTCCGGGGTGAGTT
CGCCAGGCCCTGCGGTGAGCTGGATGTGTGGTGTCTGGATGCTGCAGGTCCGGGGTGAGTTCGCCAGGCCCTCGGTGAGC
TGGATATGCGGTGTCCCGTGTCCGAATGGTGCAGGTCCAGGGTGAGGTCCGACAGGCCCTTGGTGGGCTGGATGTGCCGT
GTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGACAGGCCCTTGGTGAGCTGGATGTGCGGTGTCCGGATGGTGCAGGTCCG
GGGTGAGGTACCCAGGCCCTCGGTGATCTGGATGTGGCATGTCCTTCTCGTTTAAG

35 Intron 3 (SEQ ID NO 6)

40

GTACTGTATCCCCACGCCAGGCCTCTGCTTCTCGAAGTCTTGAACACCAGCCCGGCCTCAGCATGCGCCTGTCTCCACT
TGCCTGTGCTTCCCTGGCTGTGCAGCTCTGGGCTGGGAGCCAGGGGCCCCGTACAGGCCTGGTCCAAGTGGATTCTGTG
CAAGGCTCTGACTGCCTGGAGCTCACGTTCTCTTACTTGTAAAATCAGGAGTTGTGCCAAGTGGTCTCTAGGGTTTGT
AAGCAGAAGGGATTTAAATTAGATGGAAACACTACCACTAGCCTCCTTGCCCTTCCCTGGGATGTGGGTCTGATTCTCTC
TCTCTTTTTTTTTTCTTTTTGAGATGGAGTCTCACTCTGTTGCCAGGCTGGAGTGCAAGTGGCATAATCTGGCTCACT

GCAACCTCCACCTCCTGGGTTTAAGCGATTACCAGCCTCAGCCTCCTAAGTAGCTGGGATTACAGGCACCTGCCACCAC
GCCTGGCTAATTTTTGTACTTTTAGGAGAGACGGGGTTTACCATGTTGGCCAGGCTGGTCTCGAACTCATGACCTCAGG
TGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGTTTACAGGCTAAGCCACCGTGCCAGCCCCGATTCTCTTTAATT
CATGCTGTTCTGTATGAATCTTCAATCTATTGGATTAGGTCATGAGAGGATAAAATCCACCCACTTGGCGACTCACTG
5 CAGGGAGCACCTGTGCAGGGAGCACCTGGGGATAGGAGAGTTCCACCATGAGCTAACTTCTAGGTGGCTGCATTTGAATG
GCTGTGAGATTTGTCTGCAATGTTCCGGCTGATGAGAGTGTGAGATTGTGACAGATTCAAGCTGGATTTCATCAGTGAG
GGACGGGAGCGCTGGTCTGGGAGATGCCAGCCTGGCTGAGCCCAGGCCATGGTATTAGCTTCTCCGTGTCCCGCCAGGC
TGA CTGTGGAGGGCTTTAGTCAGAAGATCAGGGCTTCCCCAGCTCCCCTGCACTCGAGTCCCTGGGGGGCTTGTGAC
ACCCCATGCCCAAATCAGGATGTCTGCAGAGGGAGCTGGCAGCAGACCTCGTCAGAGGTAACACAGCCTCTGGGCTGGG
10 GACCCCGACGTGGTGTGGGGCCATTTCTTGCATCTGGGGAGGGTCAGGGCTTTCCTGTGGGAACAAGTTAATACAC
AATGCACCTTACTTAGACTTTACACGTATTTAATGGTGTGCGACCCAACATGGTCATTTGACCAGTATTTTGAAAGAAT
TTAATTGGGGTGACCGGAAGGAGCAGACAGACGTGGTGGTCCCCAAGATGCTCCTTGTCACTACTGGGACTGTTGTTCTG
CCTGGGGGGCTTGGAGGCCCTCCTCCCTGGACAGGGTACCGTGCTTTTCTACTCTGCTGGGCCTGCGGCTGCGGTCT
AGGGCACCAGCTCCGGAGCACCCGCGGCCCCAGTGCCACGGAGTGCCAGGCTGTGAGCCACAGATGCCAGGTCCAGGT
15 GTGGCCGCTCCAGCCCCGTGCCCCATGGGTGGTTTGGGGGAAAAGGCCAAGGGCAGAGGTGTGAGGAGCTGGTGGG
CTCATGAGAGCTGATTCTGCTCCTTGGCTGAGCTGCCCTGAGCAGCCTCTCCCGCCTCTCCATCTGAAGGATGTGGCT
CTTTCTACCTGGGGGTCTGCCTGGGGCCAGCCTTGGGCTACCCAGTGGCTGTACCAGAGGGACAGGCATCCTGTGTGG
AGGGGCATGGGTTACGTGGCCCCAGATGCAGCCTGGGACCAGGCTCCCTGGTGTGATGGTGGGACAGTACCCTGGGG
GTTGACCGCCGGACTGGGCGTCCCCAGGGTTGACTATAGGACCAGGTGTCCAGGTGCCCTGCAAGTAGAGGGGCTCTCAG
20 AGGCGTCTGGCTGGCATGGGTGGACGTGGCCCCGGGCATGGCCTTCAGCGTGTGCTGCCGTGGGTGCCCTGAGCCCTCAC
TGAGTCGGTGGGGGCTTGTGGCTTCCCGTGAAGCTTCCCCCTAGTCTGTTGTCTGGCTGAGCAAGCCTCCTGAGGGGCTCT
CTATTGCAG

Intron 4 (SEQ ID NO 7)

GTGGCTGTGCTTTGGTTTAACTTCCTTTTTAAACAGAAGTGCCTTTGAGCCCCACATTTGGTATCAGCTTAGATGAAGGG
25 CCCGGAGGAGGGGCCACGGACACAGCCAGGGCCATGGCACGGCGCAACCCATTTGTGCGCACAGTGAGGTGGCCGAGG
TGCCGGTGCCTCCAGAAAAGCAGCGTGGGGGTGTAGGGGAGCTCCTGGGGCAGGGACAGGCTCTGAGGACCAAGAAG
CAGCCGGGCCAGGGCCTGGATGCAGCACGGCCCCAGGTCTGGATCCGTGTCTGCTGTGGTGCAGCCTCCGTGCGCT
TCCGCTTACGGGGCCCGGGACCAGGCCACGACTGCCAGGAGCCACCGGGCTCTGAGGATCCTGGACCTTGCCCCACGG
30 CTCCTGCACCCACCCCTGTGGCTGCGGTGGCTGCGGTGACCCCGTCATCTGAGGAGAGTGTGGGTGAGGTGGACAGAG
GTGTGGCATGAGGATCCCGTGTGCAACACACATGCGGCCAGGAACCCGTTTCAAACAGGGTCTGAGGAAGCTGGGAGGGG
TTCTAGGTCCCGGTCTGGGTGGCTGGGGACACTGGGGAGGGGCTGCTTCTCCCTGGGTCCCTATGGTGGGTGGGCAC
TTGGCCGGATCCACTTTCCTGACTGTCTCCCATGCTGTCCCCGCCAG

Intron 5 (SEQ ID NO 8)

GTGGGTGCCGGGACCCCGTGAGCAGCCCTGCTGGACCTTGGGAGTGGCTGCCTGATTGGCACCTCATGTTGGGTGGAG
GAGGTACTCCTGGGTGGGCCGAGGGAGTGAGGTGACCCTGTCACTGTTGAGGACACACCTGGCACCTAGGGTGGAGGC
CTTCAGCCTTTCTGCAGCACATGGGGCCGACTGTGCACCTGACTGCCCGGGCTCCTATTCCCAAGGAGGGTCCCACTG
GATTCCAGTTTCCGTGAGAGAAGGAACCGCAACGGCTCAGCCACCAGGCCCCGGTGCCTTGACCCCACTGCTGAGCCAG
40 GGGTCTCCTGTCTGAGGCTCAGAGAGGGGACACAGCCCCCCTGCCCTTGGGGTCTGGAGTGGTGGGGTTCAGAGAGAG

GTGGCTGCACCTGCATCCCTGCAATCCCTCCAGCACTGGGCTGGAGAGGCCCGGGAGCTCGAGTGCCACTTGTGCCACGT
GACTGTGGATGGCAGTCGGTCACGGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTGGTCACAGGGGTCTGATGTGTG
GTGACTGTGGATGGCGGTCTGATGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGG
ATGGCGGTCTGATGTGGTGAAGTGTGGATGGCGGTCTGATGTGGTGAAGTGTGGATGGCGGTCTGAT
5 GGGTCTGATGTGGTGAAGTGTGGATGGCAGTCGGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATG
TGGTGAAGTGTGGATGGCAGTCGGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAG
GTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGG
CGGTCTGATGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGTATCGGTACAG
CAGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGTATCGGTACAG
10 GGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTGGTCCCGGGG
TCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGT
GATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGT
GACTGTGGATGGCGGTCTGATGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGAT
GGCGGTGGTCCCGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCAG
15 TCGTGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGT
TCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGT
GGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGTATCGGTACAGGGGTCTGATGTGTGGT
GACTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGAT
GGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCAGTCG
20 GTCACAGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGAT
GGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGT
GTGGTGAAGTGTGGATGGTATCGGTACAGGGGTCTGATGTGTGGTAGCTGCAGGTGGAGTCCAGGTGTGTCTGTAGCT
ACTTTCGCTCCTCGGCCCGGCCCGTTTCCCAAACAGAGCTTCCAGGCGCTCTCTGGGCTTCATCCCGCCATCG
GGCTTGGCCGACAGTCCACAGTCTGATCGGAAGAAACAAGTGCCAGCTCTGGCCGGGGCAGGCCACATTTGTGGCTC
25 ATGCCCTCTCTCTGCGGCAG

Intron 7 (SEQ ID NO 11)

GTCTGGGCACTGCCCTGCAGGGTTGGGCACGGACTCCAGCAGTGGGTCTCCCTGGGCAATCACTGGGCTCATGACCG
GACAGACTGTTGGCCCTGGGGGCGAGTGGGGGAATGAGCTGTGATGGGGCATGATGAGCTGTGTGCCTTGGCGAAATC
30 TGAGCTGGGCCATGCCAGGCTGCGACAGCTGCTGCATTACAGGCACCTGCTCACGTTTGACTGCGCGGCCTCTCTCCAGTT
CCGAGTGCCTTTGTTTCATGATTTGCTAAATGTCTTCTGCGCAGTTTGTATCTTGAGGCCAAAGGAAAGGTGTCCCCCT
CCTTTAGGAGGGCAGGCCATGTTTGAGCCGTGCTGCGCCAGCTGGCCCTCAGTGCTGGGTCTGAGGCCAAAGGAAACG
TGTCCTCTTCTTAGGAGGACGGGCCGTGTTTGAGCCACGCCCGCTGAGCGGGCCTCTCAGTGCTGGGTCTGTCCACGT
GGCCCTGTGGCCCTTTGCAGATGTGGTCTGTCCACGTGGCCCTGTGGCTCTTTGCAGATGCCTGTTAGCACTTGCTCGGC
35 TCTAGGGGACACTCGTGTCCACGCATGAGGCTCAGAGACCTCTGGGCGAATTTCTTGGCTCCAGGGTGGGGGTGGAG
GTGGCTGGGCTGCTGGGACCCAGACCCTGTGCCCCGACAGTGGGCGCAACTCCTGGATCACATATGCCATCCGGGCCA
CGGTGGGCTGTGTGGGTGTGAGCCAGCTGGACCCACAGGTGGCCACAGGAGACGTTCTGTGTACACACTCTGCCTAA
GCCATGTGTGTCTGCAGAGACTCGGCCCGGCCAGCCACAGATGGCCCTGCATTCAGCCACGCCCGCACTTCATCACA
AACACTGACCCCAAAGGGACGGAGGGTCTTGGCCACGTGGTCTGCTGTCTCAGCACCCACCGGCTCACTCCCATGTG
40 TCTCCCGTCTGCTTTCGAC

Intron 8 (SEQ ID NO 12)

GTGAGTCAGGTGGCCAGGTGCCATTGCCCTGCGGGTGGCTGGGCGGGCTGGCAGGGCTTCTGCTCACCTCTCTCCTGCCC
CTTCCCCACTGNCCTTCTGCCCCGGGGCCACCAGAGTCTCCTTTTCTGGCCCCCGCCCCCTCCGGCTCCTGGGCTGCAGGC
5 TCCCCAGGCCCCGAAACATGGCTCGGCTTGCGGCAGCCGGAGCGGAGCAGGTGCCACACGAGGCCTGGAAATGGCAAGC
GGGGTGTGGAGTTGCTCCTGCGTGGAGGACGAGGGGCGGGGGTGTGTCTGGGTGAGGTGTGCGCCGAGCGTTTGAGCCT
GCAGCTTGTGAGCTCCAAGTTACTACTGACGCTGGACACCCGGCTCTCACACGCTTGTATCTCTCTCTCCCGATACAAAA
GGATTTTATCCGATTCTCATTCCCTGTCCCTGTGCTGTGACCCCGCGAGGGCGGGGCTCTTCTCTGTGACTAGATTT
CCCATCTGGAAGTGGCGGGTTGACCGTGTAGTTGCTCCTCTCGGGGGGCTGTGGTGCCATGGGGCAGGCGGCCTGG
10 GAGAGCTGCCGTACACAGCCACTGGGTGAGCCACACTCAGGTGGTAGAGCCACAGTGCTGTGCCACATCACCTCCT
CTGGATTTTAAGTAAAACACACACCTCCCGGAGGCATCTGCCTGCGACCTGTGTGTGCTGGGGAGAGTGGTAGCAC
GGAGGAAATTCGTGCACACTCAAGGTCATCAGCAAGGTCATCCGCAGTCAGGTGGAACGTGGAGGCCTCTCTCTGGGATC
GTCTCCAGCGGATAAAGGACTGTGCACAGCTTCGGAAGCTTTTATTTAAAAATATAACTATTAATTATTGCATTATAAGT
AATCACTAATGGTATCAGCAATTATAATATTTATTAAAGTATAATTAGAAATATTAAGTAGTACACAGTTCTGAAAAA
15 CACAAATTGCACATGGCAGCAGAGTGAATTTTGGCCGAGGGACAGTGTGCACATGTGTGTAAGCGGCCCCCAGGCCCCAC
AGAATTCGCTGACAAAGTCACCTCCCCAGAGAAGCCACCACGGGCCTCCTTCGTGGTGTGTAATTTTATTAAGATGGATC
AAGTCACGTACCGTCCACGTGTGGCAGGGCTTTGGGGAATGTGAGGTGATGACTGCGTCCTCATGCCCTGACAGACAGGA
GGTGAAGTGTGTCTGTCTGTCCCTAGGACACGGACAGGCCCCGAAGCTCTAGTCCCCATCGTGGTCCAGTTTGGCCTCTGA
ATAAAACGTCCTTCAAAACCTGTTGCCCCAAAACTAAGAACAGAGAGAGTTTCCCATCCCATGTGCTCACAGGGGCGTA
20 TCTGCTTGCCTTGACTCGCTGGGCTGGCCGGACTCTAGAGTTGGTGCCTGTGCTTCTGTGCAAAAAGTGAGTCCTCTT
GCCCATCACTGTGATATCTGCACCAGCAAGGAAAGCCTCTTTCTTTTCTTTCTTTTCTTTTGTGAGACGGAACGTCA
CTGTTGTCTGCTGGGCTTGAGTGACGTGGCGGATCTCAACTCACTGCAACCTCCGCCTCCCGGGTTCCAGCATTTCTC
CTGCCCTCAGCCTCCCGAGCAGCTGAGATTACAGGCACCCACCCCTGCGCCTGGCTAATTTTGTATTTTGTAGAGAG
GGGTTTTTGGCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTG
25 GGATTACAGGTGTGAGCCATCACGCCAGCCGAAAGCCTCTTTTTAAGGTGACCACCTATAGCGCTTCCCGAAAATAAC
AGGTCTTGTTTTTGCAAGCTGCTCTTAGCAACAGGAGTGGCGTCTGTGGGCTCTGGGGATGGCTGAGGG
TCGCGTGGCAGCCATGCCTTCTGTGTGCACCTTAGGTTCCACGGGGCTATTCTGCTCTCACTGTTGTCTGAAAACGCA
CCCTTGGCATCCTTGTTTGGAGAGTTTCTGCTTCTCGTGGTCACTGAAACTAGGGGCAAGGTTGTATCCGTTGGCGC
GCAGCGGCTACATGTAGGGTCACTGAGTCTTTACCGTGGACAAATTCCTGAAAAAAGGAGTCCGGTTAAGCAT
30 TCATTCCGGGTCAAGTGTCTGGTTCTGTGAATAAACTCTAAGATTTAAGAAACCTTAATGAAAGAAAACCTTGATGATTC
AGAGCAAGGATGTGGTCACACCTGTGGCTGGATCTGTTTCAGCCGCCCCAGTGATGGTGAGAGTGGGGAGCAGGGATTG
TTGTTTCAAGAGTCTCATCTGGTATGTTTCTGAGGTGTTTCCGGGCTGAATGGTAGAGTGTGCTTTGTGTGTATGAGGT
TCTGTGTCTGTGTGGCTCGGTTTGTAGTGTACGCATGTCCAGCACATGCCCTGCCCGTCTCTCACCTGTGTCTTCCCGC
CCCAG

35

Intron 9 (SEQ ID NO 13)

GTGAGGCCTCCTCTTCCCCAGGGGGGCTTGGGTGGGGGTTGATTGCTTTTGATGCATTGAGTGTAAATATTCCTGGTGC
TCTGGAGACCATGACTGCTCTGTCTTGAGGAACCAGACAAGGTTGAGCCCCCTTCTTGGTATGAAGCCGCACGGGAGGGG
TTGCACAGCCTGAGGACTGCGGGCTCCACGCAGGCTCTGTCCAGCGGCCATGTCCAGAGGCCTCAGGGCTCAGCAGGCGG
40 GAGGGCCGCTGCCCTGCATGATGAGCATGTGAATTCAACACCGAGGAAGCACACCAGCTTCTGTACGTCACCCAGGTTCC

CGTTAGGGTCCCTTGGGGAGATGGGGCTGGTGAGCCTGAGGCCCCACATCTCCAGCAGGCCCTCGACAGGTGGCCTGGA
CTGGGCGCCTCTTCAGCCCATTGCCCATCCCCTTGATGGGGTCTACACCAAGGACGCACACCTAAATATCGTGCC
AACCTAATGTGGTTCAACTCAGCTGGCTTTTATTGACAGCAGTTACTTTTTTTTTTTAATACTTTAAGTTCTAGGGTAC
ATGTGCACGACGTGCAGGTTAGTTACATATGTATACATGTGCCATGTTGGTGTGCTGCACCCATTAACCTCATCTTTACA
5 TTAGGTATATCTCCTAATGCTATCCCTCCCCACTCCCCCATCCCATGACAGGCCCTGGTGTGTGATGTTCCCCACCCTG
TGTCGAAGTGTCTCATTGTTTCTGTTCCACCTGTGAGTGAGAACATGTGGTGTGTTGGTTTTCTTTCCTTGCAATAGTTT
GCTCAGAGTGATGGTTTCCAGCTTCGTCCATGTCCCTACAAAGGACATGAACTCATCCTTTTTTATGACTGCATAGTATT
CCGTGGTGTATATGTGCCACATTTTCTTAATCCAGTCTATCATCGATGGACATTTGGGTTGGTTGCAAGTCTTTGCTACT
GTGAATAGTGCCGCAATAAACATACGTGTGCATGTGTCTTTATAGCAGCATGATTTATAATCCTTTGGGTATATACCCAG
10 TAATGGGATGGCTGGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAGGAATCACCACACTGTCTTCCACAATGGTTGAA
CTAGTTTACACTCCCAACAGTGTAAGTGTCTGGTGCTGGAGAGGATGTGGACAGCAGTTATTTTTTTATGAAAA
TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATGCGTCAGGAAGCCTGCAGGCCACACAGCCATTTCTCTCGAAGAC
TCCGGGTTTTCTCTGTGCATCTTTTGAAACTCTAGTCCCAATTATAGCATGTACAGTGGATCAAGGTTCTTCTTCATTAA
GGTTCAAGTTCTAGATTGAAATAAGTTTATGTAACAGAAACAAAATTCTTGTAACACAACTTGCTCTGGGATTTGGA
15 GGAAAGTGTCTCGAGCTGGCGGCACACTGGTCAGCCCTCTGGGACAGGATACCTCTGGCCCATGGTCATGGGCGCTGG
GCTTGGGCGCTGAGGGTCACACAGTGACCATGCCCAGCTTCTGTGGATAGGATCTGGGTCTCGGATCATGCTGAGGACC
ACAGCTGCCATGCTGGTAAAGGGCACCAGTGGCTCAGAGGGGCGAGGTTCCAGCCCCAGCTTCTTACCCTCTTCAG
TTATTTTTCCCTAAGAGTCTGAGAAGTGGGGCGCGCCTGATGGCCTTCGTTCTTCAGCTGGCACAGAATTGCACAA
GCTGATGGTAAACACTGAGTACTTATAATGAATGAGGAATTGCTGTAGCAGTTAACTGTAGAGAGCTCGTCTGTGGA
20 GAAATTTAAGTTTTTCATTTAACCGCTTTGGAGAATGTTACTTTATTTATGGCTGTGTAAATTGTTTGACATTAGTCCC
TCGTAGACAGATACTACGTAAAAAGTGTAAGTTAACCTTGCTGTGTATTTTCCCTTATTTTAG

Intron 10 (SEQ ID NO 14)

GTGAGGCGCGTGCCGTGTGTCTGTGGGGACCTCCACAGCCTGTGGGCTTGCAGTTGAGCCCCCGTGTCTGCCCTGG
25 CACCGCAGCGTTGTCTCTGCAAGTCTCTCTCTGCGGTGCTGGATCCGCAAGAGCAGAGGCGCTTGGCCGTGCACC
CAGGCCTGGGGGCGCAGGGGCACCTTCGGGAGGAGTGGGTACCGTGACGGCCCTGGTCTGCAGAGACGCACCCAGGTT
ACACACGTGGTGAGTGACGGCGGTGACCTGGCTCCTGTGCTCTTTGGAAAGTCAAGAGTGGCGGCTCCTGGGGCCCCAG
TGAGACCCCCAGGAGCTGTGCACAGGGCCTGCAGGGCCGAGGCGGCAGCCTCCTCCCAGGGTGCACCTGAGCCTGCGGA
GAGCAGGAGCTGCTGAGTGAGCTGGCCACAGCGTTCGCTGCGGTACGTTCTGCGTGGGGTGTGTTGGGATCGGTGGG
30 AGAATTTGGATTTGCTGAGTGCTGTCTTGAACACGGAGATGGCTAGGAGTGGGTTTCAGAGTTGATTTTTGTGAAT
CAAATAAAATCAGGCACAGGGACCTGGCCTCAGCACAGGGGATTGTCCAATGTGGTCCCCCTCAAGGGCGCCCCACAG
AGCCGTTGGGCTTGTGTTTAAAGTGCATTTGACGAGGGACGAGAAACCTTGAAAGCTGTAAAGGGAACCTCAGAAAATG
TGGCCGCCAGGGGTGGTTTCAGGTGCTTTGCTGGGCTGTGTTGTGAAAACCCATTGGACCCGCCCTCAAGTCCACCC
TCCAGGTCCACCCTCCAGGGCGCCCTGGGCTGGGGTATGCTTGGCGTTCTTGTGCGCAGCCCGAGCACAGCAGGC
35 TGTGCACATTTAAATCCACTAAGATTCACTCGGGGGAGCCAGGTCCCAAGCAACTGAGGGCTCAGGAGTCTGAGGCT
GCTGAGGGGACAGAGCAGACGGGAACGCTGCTTCTGTGTGGCAAGTTCTGAGGGTGTGCGCAGGGAGGTGGCTCAGA
GTGTATGTTGGGGTCCCACGGGGGAGAACTCTGTCTCTGATGAGTCCGCAGCCATGTAAACAGGAAGGGGTGGCCACAG
GGAGCTGGGAATGCACAGGGGAGCTGCGCAGCTGGCCGAGGTCCAGGGCCAGGCCACAGGAAGGGGAGGGGGACGCC
GGGGCCACAGCAGAGGCCGAGGAAGGGAAGGGATGCCAGGCCAGAGCAGAGGCTACCGGGCAGGGGGGCTCCCTG
40 AGCTGGGTGAGCGAGGCTCATGACTCGGCGAGGGAACCTCCTTGACGTGAAGCTGACGACTGGTGTGCCCAGCTCAGAG

CCCAGCCAGGTCCCGCGCCTGAGCAGGAACTCAGAACCTCCCCCTTTGTCTAAAGCACAGCAGATGCCTTCAGGGCATCT
AGGAGAAAAACAGGCAAAGTCGTTGAGAAACGTCTTAAAAGAAGGTGGGATGGTGGCAATTTCTTGTCAGATTTTAGTCT
GCCCCGGACCACAGATGAGTCTATAACGGGATTGTGGTGTGCCATGGGGACACATGAGATGGACCATCACAGAGGCCAC
TGGGGCTGCACCTCCCATCTGAGTCTCTGGCTGTCCCGGTCCAGGCCAGGTTCTTGTCATGCTCACCTACCTGTCTGCCC
5 GGGAGACAGGGAAGCAGCCCGAAGTCTGGAGCAGGGCTGGGTCCAGGCTCCTCAGAGCTCCTGCCAGGCCAGCACCTT
GCTCCAAATCACCACTTCTCTGGGGTTTTCCAAAGCATTAAACAAGGGTGTGAGTTACCTCCTGGGTGACGGCCCCGCA
TCCTGGGGCTGACATTGCCCTCTGCCTTAG

Intron 11 (SEQ ID NO 15)

10 GTGAGCGCACCTGGCCGAAGTGAGCCTGTGCCCGGTGGGGCAGGTGCTGCTGTCAGGGCCGTTGCGTCCACCTCTGCT
TCCGTGTGGGGCAGGCGACTGCCAATCCCAAAGGGTCAGAGGCCACAGGGTGCCCTCGTCCCATCTGGGGCTGAGCAGA
AATGCATCTTTCTGTGGGAGTGAGGGTGTCTCAACGGGAGCAGTTTCTGTGCTATTTTGGTAAAAGGAAATGGTGCAC
CAGACCTGGGTGCACTGAGGTGTCTTCAGAAAGCAGTCTGGATCCGAACCCAAGACGCCCCGGGCCCTGCTGGGCGTGAGT
CTCTCAAACCCGAACACAGGGCCCTGCTGGGCATGAGTCCCTCTGAACCCGAGACCCTGGGGCCCTGCTGGGCGTGAGT
15 CTCTCCGAACCCAGAGACTTCAGGGCCCTTTGGGCGTGAGTCTCTCCGCTGTGAGCCCCACACTCCAAGGCTCATCCAC
AGTCTACAGGATGCCATGAGTTCATGATCACGTGTGACCCATCAGGGGACAGGGCCATGGTGTGGGGGGGTCTCTACAA
AATTCTGGGGTCTTGTTCCTCCAGAGCCCGAGAGCTCAAGGCCCGTCTCAGGCTCAGACACAAATGAATTGAAGATGGA
CACAGATGCAGAAATCTGTGCTGTTCTTTTATGAATAAAAAGTATCAACATTCCAGGCAGGGCAAGGTGGCTCACACCT
ATAATCCCAGCACTTTGGGAGGCCGAGGTGGGTGGATCACTTGAGGCCAGGAGTTTGAGGCCAACCTAACCAACATAGTG
20 AAATTCATTTCTACTTAAAAAATACAAAAATTAGCCTGGCCTGGTGACACGCCTGTAGTCCCCGCTATGCGGGAGGC
TGAGGCAGGAGAATCATTTGAACCCAGGAGGCAGAGGTGTCAGTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCA
ACAGAGTGAGACTTCATCTTAAAAAAGAGTATCAGCATTCCAAAACCATAGTGACAGGTGTTTTTTTATTCT
TGTCCTTCGATAATATTTACTGGTGCTGTGCTAGAGGCCGAACTGGGGGTGCCTTCCTCTGAAAGGCACACCTTCATGG
GAAGAGAAATAAGTGGTGAATGGTTGTTAAACCAGAGGTTTAACTGGGGTCTGTGCTTCTGAGTTAACAGTCCAGATC
25 TGGACTTTGCCTCTTCCAGAAATGCTCCCTGGGGTTTGCTTCATGGGGGAGCAGCAGGTGTGGACACCCTCGTGATGGGG
GAGCAGCAGGTGCAGACGCCCTCATGATGGGGGAGTGGCAGGTGCAGACACCCTTGTCATGGTGGCCAGCATGTCCCTG
TTGCAGCTCCCTCCCCACAAGGATGCCGTCTCCTGTGCTCCCCACAGTCCCTGCTTCCCTCTCACAGCCTTACCTGGTC
CTGGCTCCACTGGCTTTGTCTGCATGATTCCACATTTCTGGGCTCCCAGCACCTCTTCGCCTCTCCAGGCACCTCT
GCAGTGCTGGCCATACCAGTCAGCTGTGAAGTGTCCACTGCTTATTTTGCTCCCCATGAAATGTATTTTTTAGGACAGGC
30 ACCCTGGTTCCAGCCTCTGGCACAGCATCAGTGAATGTTATTGAAGGACAAAGGACAGACAAACAAATCAGGAAAATGG
GTTCTCTCTAAACACATTGCAAAGCCACAGAGGCTAGTGAGGATGGGTGGGCATCAGGTTCATCAGATGTGGTCCAATG
CCAGAATATTCTGTGCTCCCAAAGGCCACTTGGTCAGAGTGTGTGCTTGACAGAGGTGGCTCTAAAAGCTCAGCAGTGGAG
GCAGTGGTTCGCCATACTCAGGGTGAACTCACATCCTCTGTGTCTGAAGTATACAGCAGAGGCTTGAAGGGCATCTGGGA
GAAGAAAACAGGCAAAATGATTAAGAAAAGTGAAAAGGAAAAGTGGTAAGATGGGAATTTTCTTGTCAGATTTTAGTCT
35 TCCCAAACACAGCTCAGATGGTAGAATGTGGTCAGAACTGATGGACAGAACAATAGAACAACCGGAAGCCCTATCTCT
CAGAAACGTGTGTTAATGTGGTATGTGGCACAGCTGATGAAAAGAGAGTGTGTGTGAATTTTTTTTCTGAGAAAAC
GACTGGAAGCAAATAAGTTGTGTCTTTACAGCATATACCAGAGCAGATTCTAGGTAGAAGAGGAGACATGCAAACAAC
ACCAGCAACAGAAATAAAACAAAAGACTCAAAGGGAAGGAGGTGAACGTTCCCTGGTTGGTGTGGGGAAGGACACAC
AGGGAGGCGGATGAAACAGTGAGGCAACGGGCATTGCTTTCACTGCAGAGAACTCAGCTTGCTGAGCCACAGTGAAA
40 ATGGCCATTCCCTGGAGCGTTTGTGCACGTGATTTATTTAAGGCGCCCTGTGAGGTCTGCACATTATCCTCTCACTTT

5 GTTCTCCTAACCACCTGAGAGGTAGAGGAGGAAAGGCTCCAGGGGAGCAGCCGCCCTTGGTCACCCAGCTGGCAAAGGGC
ATGCATGATTGCAGCCTGGCCTCCTGCTCCGGGGCCCTTGCTCTGCCGAGGACCCACACAAGTCAGACCCATAGGCTC
AGGGTGAGCCGGAGCCCAAGGTCGTGTTGGGGATGGCTGTGAAAGAAGAAATGGACGTCTGATGCACACTTGGGAAGGTC
CTACCAGCAGCGTCAAAGAAATGCATGTGAAACTGACAGCGAGACCCATCCCTCAAAGAAACGCACGTGAAACTGATGGC
10 GAGACCTGTCCCATCCCTCATGCTGGCTCCTTTTCTGGGCTTGCCAAGAGCCAGCATCAGGTTGAGGCAAGCTGGAAAG
ACTTTTCTGGAAAGCAGCTTGTGTCATGGAAGTCCTCACAATGTCTGTCTTCCCAGTAATTCCACTTCTGAAGTGA
CCAGACATTATCACGGGTCTTATTTACCATTTCAGTGTTCCAGGCAGGGGGACTTGCCACAGCAAGTCACGAACCTGCC
CAAATACAGGGCTAAGGAGATATTATGCATCACAAACTTGCTCTGCCATTAAACATTTTTCAAAGAATTTTGAAGAAT
GTTTAATGGCACAAAACGTTTATTTCAATGTAGCAGTGTTCAAAGCTGGATGTAAAAGAACACACCCAGGAGCCTGCCG
15 TGAATGTCATGTGTGTTTCTTTGGACATGGACATACATGGGCAGTGAGTGGTGGTGAGGCCCTGGAGGACATCGGTGG
GATGCCCTCCATCCTGCCCCCTGGAGACACCATGTGTGCCAGTGCACTCACTGGAGCCCTGTTTAGCTGGTGCCACCTG
GCTCTTCCATCCCTGAGATTCAAACACAGTGAGATTCCCCACGCCCAACTCAGTGTTCTCCACAAAAAACCTGAGTCAC
ACCTGTGTTCACTCGAGGGACGCCCGGGAGCCAGGGCTCCACAGTTTATTATGTGTTTTGGCTGAGTTATGTGCAGATC
TCATCAGGGCAGATGATGAGTGACAAACACGGCCGTGCGAGGTTTGGATACACTCAACATCACTAGCCAGGTCTCTGGTG
20 GAGTTTGGTCATGCAGAGTCTGGATGGCATGTAGCATTGGAGTCCATGGAGTGAGCACCAGCCCCCTCGGGCTGCAGC
GCATGCCCCAGGCAGGACAAGGAAGCGGGAGGAAGGCAGGAGGCTCTTTGGAGCAAGCTTTGCAGGAGGGGGCTGGGTGT
GGGCGAGGCACCTGTGTCTGACATTCCCCCTGTGTCTCAG

Intron 12 (SEQ ID NO 16)

20 GTGAGCAGGCTGATGGTCAGCACAGAGTTCAGAGTTCAGGAGGTGTGTGCGCAAGTATGTGTGTGTGTGTGCGCGCT
GCCTGCAAGGCTGATGGTGACTGGCTGCACGTAAGAGTGACATGTACGCATATACAGTGAGCACATACATGTGTGCAT
GTGTGTACATGAAGGCATGGCAGTGTGTGCACAGGTGTGCAAGGGCACAAGTGTGTGCACATGCGAATGCACACCTGACA
TGATGTGTGTTTCGTGCACAGTCGTGTGGGCATTACAGTGAGGTGCATGCGTGTGGGTGTGAGTGTAGCATGTGT
GCACATAACATGTATTGAGGGGTCCTCGTGTTCACCCCGCTAGGTCTCAGCACCAGTGCCACTCCTTACAGGATGAGAC
25 GGGGTCCCAGGCCTTGGTGGGCTGAGGCTCTGAAGCTGCAGCCCTGAGGGCATTGTCCCATCTGGGCATCCGCGTCCACT
CCCTCTCCTGTGGGCTTCTGTGTCCACTCCCCCTCTCCTGTGGGCATTTACATCCACTCCACTCCCTCTCTCCTGTGGGC
ATCCGCGTCCACTCCCCCTCTCTGTGGGCATCTGCGTCCACCTCCCCCTCTCTGTGGGCATTGCGTCCACTCCCTCTCCT
GGTTCCTTCTGTCTTGGCCGAGCCTCGGGGGCAGGCAGATGACACAGAGTCTTGACTCGCCAGGGTGGTTTCGAGCTG
CCGGGTGAGGGCCAGGCCGATTCTACTGGGAAGAGGGATAGTTTCTGTCAAATGTTCTCTTTCTGTTCCATCTGA
30 ATGGATGATAAAGCAAAAAGTAAAACTTAAATCCAGAGAGGTTTCTACCGTTTCTCACTCTTTCTTGGCGACTCTAG

Intron 13 (SEQ ID NO 17)

35 GTGAGCCGCCACCAAGGGGTGCAGGCCAGCCTCCAGGGACCCTCCGCGCTCTGCTCACCTCTGACCCGGGGCTTCACCT
TGGAACCTCTGGGTTTTAGGGGCAAGGAATGTCTTACGTTTTTCAGTGGTGCTGCTGCCTGTGCACAGTTCTGTTTCGCGTG
GCTCTGTGCAAAGCACCTGTTCTCCATCTCTGGGTAGTGGTAGGAGCCGGTGTGGCCCCAGGTGTCCCCACTGTGCCTGT
GCACTGGCCGTGGGACGTATGGAGGCCATCCAGGGCAGCAGGGGCATGGGGTAAAGAGATGTTTATGGGGAGTCTTAG
CAGAGGAGGCTGGGAAGGTGTCTGAACAGTAGATGGGAGATCAGATGCCCGAGGATTGGGGTCTCAGCAAAGAGGGCC
GAGGTGGGTGCAGGTGAGGGTCGCTGGCCCCACCCCGGAAGGTGCAGCAGAGCTGTGGTCCCCACACAGCCCGGCCA
GCACCTGTGCTCTGGGCATGGCTGTGCTCCTGGAACGTTCCTGTCTGGTGGTCAAGGGGTGCCCTGCCAAGAATCG
40 ACAACTTTATCACAGAGGAAGGGCCAATCTGTGGAGGCCACAGGGCCAGCTTCTGCTGGAGTCAGGGCAGGTGGTGGC

ACAAGCCTCGGGGCTGTACCAAAGGGCAGTCGGGCACCACAGGCCCGGGCCTCCACCTCAACAGGCCTCCGAGCCACTG
GGAGCTGAATGCCAGGAGGCCGAAGCCCTCGCCCCATGAGGGCTGAGAAGGAGTGTGAGCATTGTGTATTACCCAGGGCCG
AGGCTGCGCGAATTACCGTGCACACTTGATGTGAAATGAGGTCGTCTATCGTGGAAACCCAGCAAGGGCTCACGGGA
GAGTTTTCCATTACAAGGTCGTACCATGAAAATGGTTTTTAACCCGAGTGCTTGCCTTCATGCTCTGGCAGGGAGGGC
5 AGAGCCACAGCTGCATGTTACCGCCTTTGACCAGCTCCAGAGGCTTGGGACCAGGCTGTCTCAGTTCAGGGTGCGTCC
GGCTCAGACCGCCCTCCTCTCTGCCTTCTCTCTCTGCCTCAAATCTTCCCTCGTTTGATCTCCCTGACGCGTGCCTGGG
CCCTCGTGCAAGCTGCTTGACTCCTTTCCGGAAACCTTGGGGTGTGTGGATACAGGTGCCACTGAGGACTGGAGGTGT
CTGACACTGTGGTTGACCCAGGGTCCAGCTGGCGTGTCTGGGGCCTCCTTGGGCCATGATGAGGTGAGAGGAGTTTTCC
CAGGTGAAAACCTCCTGGGAACTCCAGGGCCATGTGACCTGCCACCTGCTCCTCCCATATTAGCTCAGTCTTGTCTC
10 ATTTCCCCACAGGGTCTCTAGCTCCGAGGAGCTCCCGTAGAGGGCTTGGGCTCAGGGCAGGGCGGCTGAGTTTCCCCAC
CCATGTGGGGACCTTGGGTAGTCGCTTGATTGGGTAGCCCTGAGGAGGCCGAGATGCGATGGGCCACGGGCCGTTTCCA
AACACAGAGTCAGGCACGTGGAAGGCCAGGAATCCCTTCCCTCGAGGCAGGAGTGGGAGAACGGAGAGCTGGGCCCCG
ATTTACGGCAGCCAGGCTGCAGTGGGCGAGGCTGTGGTGGTCCACGTGGCGCTGGGGGCGGGTCTGATTCAAATCCGC
TGGGGCTCGGCCTTCTGCCCCGTGTGGCCGCGCTCCACACGGGCTTGGGGTGGACGCCCCGACCTCTAGCAGGTGGC
15 TATTTCTCCCTTTGGAAGAGAGCCCCCTACCCATGCTAGGTGTTTCCCTCCTGGGTGAGAGCGTGGCCGTGTGGCAACC
CCGGGACCTTAGGCTTATTTATTTGTTTAAAAACATTCTGGGCCTGGCTTCCGTTGTGCTAAATGGGGAAAAGACATCC
CACCTCAGCAGAGTTACTGAGAGGCTGAAACCGGGGTGCTGGCTTGACTGGTGTGATCTCAGGTCAATCCAGAAGTGGCT
CAGGAAGTCAGTGAGACCAGGTACATGGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGGCTCAGGCAGTGGGT
GAGGCCAGGTACATGGGGGGCTCAGGCACTGGGTGAGATGAGGTACACGGGGGGCTCAGGCAGAGGGTCAGACCAGGTAC
20 ACGGGGGCTCTGATCACACGCACATATGAGCACATGTGCACATGTGTGTTTCATGGTAGCCAGGTCTGTGCACACCTGC
CCCAAAGTCCCAGGAAGCTGAGAGGCCAAAGATGGAGGCTGACAGGGCTGGCGCGGTGGCTCACACCTGTAGTCCCAGCA
CTTTGGGAGGCCGAGGCGAGAGGATCCCTTGAGCCCAGGAGTTTAAGACCAGCCTGAGCAACATAGTAGAACCCCATCTC
TATGAAAAATAAAAAACAAAATTAGCTGAACATGGTGGTGTGCGCCTGTAGTTCCAATACTTGGGAGGCTGAAGTGGGAG
GATCACTTGAGCCCAGGAGGTGGAAGCTGCAGTGAGCTGAGATTGCACCACTGTACTGCAGCCTGGGTGACAGAGTGAGA
25 GCCCATCTCAACAACAACAAAGAAGACTGACAAATGCAGTTTCTTGAAAGAAACATTTAGTAGGAACTTAACCTACACA
CAGAAGCCAAGTCGGTGTCTCGGTGTCTCAGTGAGATGAGATGATGGGTCTCACACCATCACCCAGACCCAGGGTTTATG
CACCACAGGGGCGGGTGGCTCAGAAGGGATGCGCAGGACGTTGATATACGATGACATCAAGTTGTCTGACGAAGGGCAG
GATTCATGATAAGTACCTGCTGGTACACAAGGAACAATGGATAAACTGGAAACCTTAGAGGCCTTCCCGGAACAGGGGCT
AATCAGAAGCCAGCATGGGGGGCTGGCATCCAGGATGGAGCTGCTTCAGCCTCCACATGCGTGTTCATACAGATGGTGCA
30 CAGAAACGCAGTGATCCTGTGCACACACAGACACGCAGCTACTCGCACACACAAGCACACACAGACATGCATGCATGC
ATCCGTGTGTGTGCACCTGTGCCATGAGGAAACCATGCATGTGCATTATGCACGCACACAGGCACCGGTGGGCCAT
GCCACACCCACGAGACCGTCTGATTAGGAGGCCTTTCCTCTGACGCTGTCCGCCATCCTCTCAG

Intron 14 (SEQ ID NO 18)

35 GTATGTGCAGGTGCCTGGCCTCAGTGGCAGCAGTGCCTGCTGGTGTAGTGTGTCAGGAGACTGAGTGAATCTGGG
CTTAGGAAGTTCTTACCCCTTTTCGCATCAGGAAGTGGTTTAACCAACCACTGTCAGGCTCGTCTGCCCCCCTCTCGT
GGGGTGAGCAGAGCACCTGATGGAAGGGACAGGAGCTGTCTGGGAGCTGCCATCCTTCCCACCTTGCTCTGCCTGGGGAA
GCGCTGGGGGGCCTGGTCTCTCTGTTTGGCCATGGTGGGATTGGGGGGCCTGGCCTCTCTGTTTGGCCTGTGGTGG
GATTGGGCTGTCTCCCGTCCATGGCACTTAGGGCCCTGTGCAAACCCAGGCCAAGGGCTTAGGAGGAGGCCAGGCCAG
40 GCTACCCACCCCTCTCAGGAGCAGAGGCCGCGTATCACACGACAGACCCCGCGCGTCTCTGCTTCCAGTCAACG

TCCTCTGCCCCCTGGACACTTTGTCCAGCATCAGGGAGGTTTCTGATCCGTCTGAAATTCAAGCCATGTCGAACCTGCGGT
CCTGAGCTTAACAGCTTCTACTTTCTGTTCTTTCTGTGTTGTGGAAATTTACCTGGAGAAGCCGAAGAAAACATTTCTG
TCGTGACTCCTGCGGTGCTTGGGTGCGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTCCGGGTGTGGGCAGCTTTCCG
GTGTCTCCTGGGAGGGGAGCTGGGCTGGGCCTGTGACTCCTCAGCCTCTGTTTTCCCCCAG

5

Intron 15 (WEQ ID NO 19)

GCAAGTGTGGGTGGAGGCCAGTGCAGGGCCCCACCTGCCAGGGGTATCCTTGAACGCCCTGTGTGGGGCGAGCAGCCTC
AGATGCTGCTGAAGTGACAGCGCCCCCGGCTGACCCTGGGGGCTGGAGCCAGCTGGCAGCCCTATGTGATTAAACG
CTGGTGTCCCCAGGCCACGGAGCCTGGCAGGGTCCCCAACTTCTTGAACCCCTGCTTCCCATCTCAGGGGCGATGGCTCC
CCACGCTTGGGAGCCTTCTGACCCCTGACCTGTGTCTCTCACAGCCTCTTCCCTGGCTGCTGCCCTGAGCTCCTGGGGT
CCTGAGCAAGTTCTCTCCCCGCCCCGCGCTCCAGCGTCACTGGGCTGCCTGTCTGCTCGCCCCGGTGGAGGGGTGTCTG
TCCCTTCACTGAGGTTCCACCAGCCAGGGCCACGAGGTGCAGGCCCTGCCTGCCCGGCCACCCACACGTCTTAGGAGGG
TTGGAGGATGCCACCTCTGGCCTCTTCTGGAACGGAGTCTGATTTTGGCCCCGCAG

10

15

3'-untranskribierter Bereich (SEQ ID NO 20)

ATCTCATGTTTGAATCCTAATGTGCACTGCATAGACACCACTGTATGCAATTACAGAAGCCTGTGAGTGAACGGGGTGGT
GGTCAGTGCAGGGCCCATGGCCTGGCTGTGCATTTACGGAAGTCTATGAGTGAATGGGGTTGTGGTCAGTGCAGGGCCCATG
GCCTGGCTGGGCCTGGGAGGTTTCTGATGCTGTGAGGCAGGAGGGGAAGGAGGGTAGGGGATAGACAGTGGGAGCCCCCA
CCCTGGAAGACATAACAGTAAGTCCAGGCCCGAAGGGCAGCAGGGATGCTGGGGGCCAGCTTGGGGCGCGGGGATGATG
GAGGGCCTGGCCAGGGTGGCAGGGATGATGGGGGCCAGCTGGGGTGGCAGGGGTGATGGGGGGGGTGGTCTGGGTGG
GGGGGAAGATGGGGAAGCCTGGCTGGGCCCCCTCCTCCCTGCCTCCACCTGCAGCCGTGGATCCGGATGTGCTTCCCT
GGTGACATCCTCTGGGCCATCAGCTTTCATGGAGGTGGGGGGCAGGGGCATGACACCATCCTGTATAAAATCCAGGATT
CCTCCTCCTGAACGCCCAACTCAGGTTGAAAGTCAATTCCGCCTCTGGCCATTCTCTTAAGAGTAGACCAGGATTCTG
ATCTCTGAAGGGTGGGTAGGGTGGGGCAGTGGAGGGTGTGGACACAGGAGGCTTCAGGGTGGGGCTGGTATGCTCTCTC
ATCCTCTTATCATCTCCCAGTCTCATCTCTCATCCTCTTATCATCTCCCAGTCTCATCTGTCTTCTCTTATCTCCCAGT
CTCATCTGTCTATCCTCTTACCATCTCCCAGTCTCATCTCTTATCCTCTTATCTCCTAGTCTCATCCAGACTTACCTCCCA
GGGCGGGTGCCAGGCTCGCAGTGGAGCTGGACATACGTCTTCTCAGGCAGAAGGAAGTGAAGGATTGCAGAGAACAG
GAGGGGCGGCTCAGAGGGACGAGTCTTGGGGTGAAGAAACAGCCCCCTCCTCAGAAGTTGGCTTGGGCCACACGAAACCG
AGGGCCCTGCGTGAGTGGCTCCAGAGCCTTCCAGCAGGTCCCTGGTGGGGCCTTATGGTATGGCCGGGTCTACTGAGTG
CACCTTGGACAGGGCTTCTGGTTTGTAGTGACGCCGGACGTGCTGGTGTGCGGGTGGGGCTTATGGCCACTGGATATG
GCGTCATTTATTGCTGCTGCTTCAGAGAATGTCTGAGTGACCGAGCCTAATGTGTATGGTGGGGCCCAAGTCCACAGACTG
TGTCGTAAATGCACTCTGGTGCCTGGAGCCCCGTATAGGAGCTGTGAGGAAGGAGGGGCTTGGCAGCCGGCCTGGGG
GCGCCTTGGCCCTGCAAACTGGAAGGGAGCGGCCCGGGCGCGTGGGGCGGACGACCTCAAGTGAGAGGTTGGACAGAAC
AGGGCGGGGACTTCCCAGGAGCAGAGGCCGCTGCTCAGGCACACCTGGGTTTGAATCACAGACCAACaGGTCAGGCCATT
GTTCACTATCCATCTTCTACAAAGCTCCAGATTCTGTTTCTCCGGGTGTTTTTGTGTAATTTTACTCAGGATTACT
TATATTTTTTGTAAAGTATTAGACCCCTTAAAAAAGGTATTTGCTTTGATATGGCTTAAGTCACTAAGCACCTACTTTAT
TTGCTGTTTTTATTATTATTATTATTATTATTAGAGATGGTGTCTACTCTGTCAACCCAGGTTGTTAGTGACAGTGGCAC
AGTCATGGCTCGCTGTAGCCGCAACCCCCAGGCTCAAGTGATCCTCCGGCCTCAGCTTCCCAGAGTGCTGGGATTACAG
GTGTGAGCCACTGCCCTTGCTGGCACTTTTAAAAAACCCTATGTAAGGTGAGGTCCAGTGCGCTTCCACACCTGTATCC
CAGTAGTTTGGGAAGCCGAGGCAGAAGGATTGTCTGAGGCCAGGAGTTTGAAGACCAGCATGGGTAACATAGGGAGACCCC

20

25

30

35

40

ATCTCTACAAAAAATGCAAAAAGTTATCCGGGCGTGGGGTCCAGCATCTGTAGTCCCAGCTGCTCGGGAGGCTGAGTGGG
AGGATCGCTTGAGCCCGGGAGGTCATGGCTGCAGTGAGCTGTGATTGTACCATCGCACTCCAGCCTGGGCAACAGAGTGA
GACCTGTCTCAAAAAAAAAAAAAAAAAAGAAGGAGAAGGAGAAGAGAAGAAGGAAGAAGGAAAGAGAAGAAGAAG
5 GAAGAAGGAAGAAAGAAGGAGAAGGAGGCCTGCTAGGTGCTAGGTAGACTGTCAAATCTCAGAGCAAAATGAAAAATAACA
AAGTTTTAAAGGGAAAGAAAAACCCAGCTCTTTGGAATTCCTTAGGCCTGAACTTCATCTCAAGCAGCTTCCTTCCACA
GACAAGCGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGGAGGAGAAGCAGGCAAGGGTGGAGGCTGTGGGTGACACCA
GCCAGGACCCCTGAAAGGGAGTGGTTGTTTTCTGCTCAGCCCCACGCTCCTGCCGGTCTGCACCTGCTGTAACCGTC
GATGTTGGTGCCAGGTGCCCACCTGGGAAGGATGCTGTGCAGGGGGCTTGCCAACTTTGGTGGGTTTCAGAAGCCCCAG
10 GCACTGTGGCAGGCACAATTACAGCCCTCCCCAAAGATGCCACGTCCTTCTCCTGGAACCTGTGAATGTGTACCCG
CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGGAATCACGGCTGCCAGTCAGCCGATCTTAAGGTATCTGATTATCTGG
TGGGCCTGATATGGCCACAAGGGTCCCTAGAAGTGAGAGAGGGAGGCAGGGGAGAGTCAGAGAGGGGACGTGAGAAGGAC
CACTGGCCACTGCTGGCTTTGAGATGGAGGAGGGGTCCCCAGCCAAGGAATGGGGGCAGCCGCTCCATGCTGGAAAAGC
AAGCAATCCTCCCCGGTCTGAGGGCACACGGCCCTGCCACGCTCGATTTTCAGGCCAGTGGGACCTGTTTCAGCTTTC
CGGCCTCCAGAGCTGTAAGATGATGCGTTTGTGTTTCAGCCACTAAGCTGCAGTGATTCGTACAGCAGCAATGGAATAG
15 CAGTACAGGGAAATGAATACAGGGACAGTTCAGAGTGACTCTCAGCCACCCCTGGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

10

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

15

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorengeht. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des offenen Leserahmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

20

25

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative, 3'-lokalisierte Splicestelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

30

Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, läßt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplitt wird.

Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer internen 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein veränderter C-Terminus resultiert.

Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätzlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind. Darüberhinaus wurde im Intron 1 eine Spl-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

Beispiel 6

5

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA⁺-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur
10 Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv
markierter Primer (5'-GTTAAGTTGTAGCTTACACTGGTTCTC 3'; 2,5-8x10⁵
cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl
inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL),
2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5 µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl
15 ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser
erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die
Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach
Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden
2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150 µl
20 Phenol/Chloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-
Acetat und 450 µl Ethanol für 45 min bei -70°C gefällt und anschließend für 15 min
bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol
gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stopplösung gelöst. Nach 5
min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel
25 aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom
ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in
Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart (TTA₁TTGT)
30 repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv (PyPyA_nNa/tPyPy) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

5 In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-losten Promotoren (Smale, 1997) einzuordnen ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei
10 anderen TATA-losten Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgene (Wick *et al.*, 1995).

Beispiel 7

15 Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp -60 bis -105 eingegrenzt.

20 Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit“ (Clontech) nach Angaben der Hersteller synthetisiert. In einem
25 Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer
30 zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer

angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer C5S: 5'-CTGCGTCCTGCTGCGCACGTGGGAAGC-3', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGGCGCGAGTTTCAGGCAG-3', 5'-flankierende Region -74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCCTTCCTTTCC-3', 5'-flankierende Region -112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCTCCGCGC-3', 5'-flankierende Region -191 - -171; Primer RP-3A: 5'-CTAGGCCGATTTCGACCTCTCTCC-3', 5'-flankierende Region -427 bis -405) mit dem 3'-PCR Primer C5Rück (5'-GTCCCAGGGCACGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprimten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rück, C5S-C5Rück und PRO-TEST1-C5Rück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

Beispiel 8

In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp - 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp -3872 bis bp -3113 bzw. bp -5363 bis bp -3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

Die Suche nach möglichen Transkriptionsfaktorbindungstellen wurde mit dem „Find Pattern“-Algorithmen aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

Translations-Startcodon ATG indentifiziert: fünf Sp1-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen -1788 bzw. -3995 der 5'-flankierenden Region gefunden.

5

Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschieden lange hTC-Promotorsequenzabschnitte generiert und 5' vor das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACGCAGCGCTGCCTGAACTCG-3', Position -43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-CCAGATCTCTGGAACACAGAGTGGCAGTTTCC-3', Position -4093 bis -4070) wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaares PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGTGGGGATTTCAG-3', Position -3120 bis -3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primerkombination PK-3A und PK-5D (5'-

GGAGATCTGATCTTGGCTTACTGCAGCCTCTG-3', Position -2110 bis -2087) amplifiziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGTCTGGATTCTGGGAAGTCCTCA-3', Position -1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmentes (NPK27). Der PK-3A Primer enthält eine HindIII Erkennungssequenz. Die
5 verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIAquick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und
10 anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5 α -Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für
15 die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.

Beispiel 10

20 Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection
25 bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

- 5 CHO-K1 und HEK 293 Zellen wurden in wassergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO₂ kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH₂PO₄ pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit
10 einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszu-plattieren.

- Für die Transfektion wurden pro Well jeweils 2×10^5 HEK 293 Zellen in einer 24-
well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium
15 entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV β-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serum-freies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium
20 hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24 Stunden wurden die Zellen zur Bestimmung der Luziferase- und der β-Gal-Aktivität geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8] mit H₃PO₄; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15 Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl
25 Luziferase-Assaypuffer (20 mM Tricin; 1,07 mM (MgCO₃)₄ Mg(OH)₂·5H₂O; 2,67 mM MgSO₄; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM Luciferin, 530 µM ATP) gemischt und das durch die Luziferase generierte Licht gemessen.

- 30 Zur Messung der β-Galaktosidaseaktivität wurden gleiche Mengen Zellysats und β-Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM MgCl₂;

50 mM β -Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100 μ l 1 M Na_2CO_3 gestoppt und die Absorption bei 420 nm bestimmt.

- 5 Für die Analyse des hTC-Promotors wurden vier verschieden lange hTC-Promotorsequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK
10 293 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstrukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im
15 Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstrukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

Literaturverzeichnis

- 5 Allsopp, R. C., Vazire, H., Pattersson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W. und Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci. 89, 10114-10118.
- 10 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). Current protocols in molecular biology. Greene Publishing Associates and Wiley-Intersciences, New York.
- Blasco, M. A., Rizen, M., Greider, C. W. und Hanahan, D. (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. Nature Genetics 12, 200-204.
- 15 Broccoli, D., Young, J. W. und deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. Proc. Natl. Acad. Sci. 92, 9082-9086.
- Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 11, 1921-1929.
- 20 Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B. (1995). The RNA component of human telomerase. Science 269, 1236-1241.
- 25 Geng, Y., and Johnson, L.F. (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthase promoter. Mol. Cell. Biol 14:4894.
- 30 Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. Science 249, 1129-1133.
- 35 Harley, C.B., Futcher, A.B., Greider, C.W., 1990. Telomeres shorten during ageing of human fibroblasts. Nature 345, 458-460.

- Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 346, 866-868.
- 5 Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* 155, 3711-3715.
- 10 Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011-2015.
- Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.
- 15 Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276: 561-567.
- Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* 57, 633-643.
- 20 McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26, 234-282.
- 25 Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc. Natl. Acad. Sci.* 86, 7049-7053.
- Olovnikov, A. M. (1973). A theory of marginotomy. *J. Theor. Biol.* 41, 181-190.
- 30 Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell* 75, 729-739.
- Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* 15, 7155-7174.
- 35 Smale, S.T. and Baltimore, D. (1989). The „initiator“ as a transcription control element. *Cell* 57:103-113.

- Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* 1351, 73-88.
- 5 Shay, J. W. (1997). Telomerae and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.
- Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C.B. und Landsdorp, P.M. (1994). Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with age. *Proc. Natl. Acad. Sci.* 91, 9857-9860.
- 10 Wick, M., Härönen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995). Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. *Biochemical Journal* 311, 549-554.
- 15 Zakian, V. A. (1995). Telomeres: Beginning to understand the end. *Science* 270, 1601-1607.

Patentansprüche

1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen
Telomerase-Untereinheit.
5
2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um
Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser
Sequenzen handelt.
10
3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um
die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen
katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder
um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.
15
4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der
Ansprüche 1 bis 3.
5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß
20 es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide
oder Proteine kodieren.
6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.
- 25 7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem
der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.
8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vek-
toren gemäß einem der Ansprüche 4 bis 6.
30

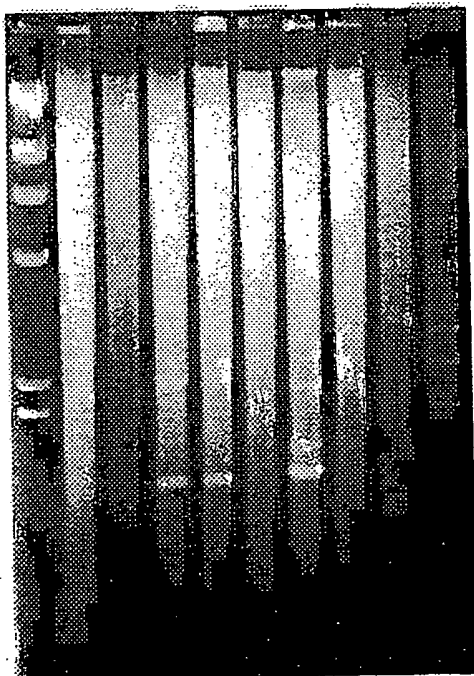
9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:
- 5 A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,
- B. Messung des Substanzeffektes auf die Reportergenexpression.
- 10
10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screenet.
- 15
11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.
- 20
12. Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:
- A. Inkubation eines rekombinanten Konstruktes bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält
- 25 mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

- C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reporter-genkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.

Fig. 1

A

1 2 3 4 5 6 7 8 9 10



B

1 2 3 4 5 6 7 8 9 10

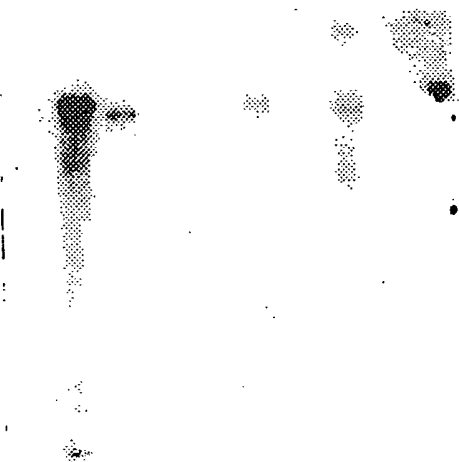


Fig. 2

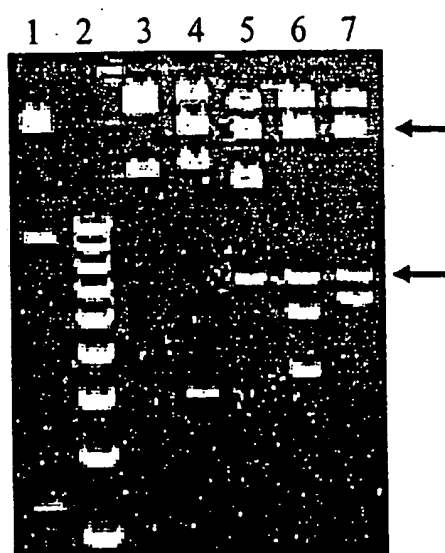
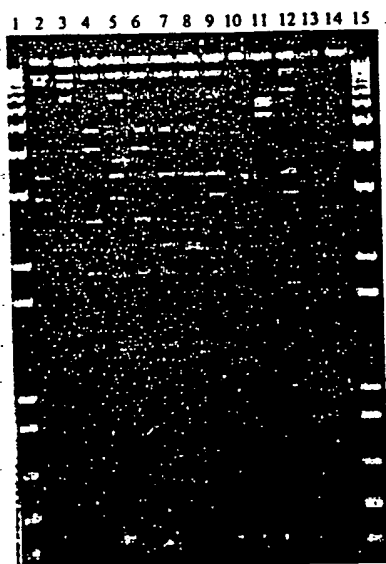


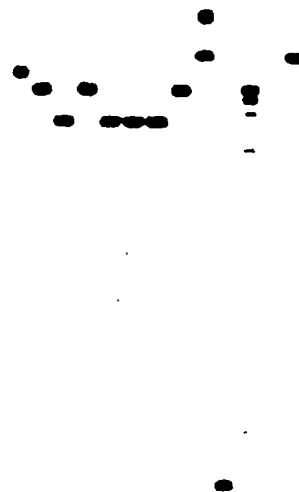
Fig. 3

A



B

2 3 4 5 6 7 8 9 10 11 12 13 14



| | | | | | | | |
|-------------|-------------|-------------|-------------|-------------|------------|-------------|------|
| GAGCTCTGAA | CCGTGGAAAC | GAACATGACC | CTTGCCCTGCC | TGCTTCCCTG | GGTGGGTCAA | GGGTAATGAA | 70 |
| GTGGTGTGCA | GGAAATGGCC | ATGTAAATTA | CACGACTCTG | CTGATGGGGA | CCGTTCCCTC | CATCATTATT | 140 |
| CATCTTCCACC | CCCAAGGACT | GAATGATTC | AGCAACTTCT | TCGGGTGTGA | CAAGCCATGA | CAAAACTCAG | 210 |
| TACAAACACC | ACTCTTTTAC | TAGGCCACACA | GAGCACGGGC | CACACCCCTG | ATATATTAAG | AGTCCAGGAG | 280 |
| AGATGAGGCT | GCTTTCAGCC | ACCAGGCTGG | GGTGACAACA | GCGGCTGAAC | AGTCTGTTCC | TCTAGACTAG | 350 |
| TAGACCCTGG | CAGGCACTCC | CCCAAATTCT | AGGGCCTGGT | TGCTGCTTCC | CGAGGGCGCC | ATCTGCCCTG | 420 |
| GAGACTCAGC | CTGGGGTGCC | ACACTGAGGC | CAGCCCTGTC | TCCACACCTT | CCGCTCCAG | GCCTCAGCTT | 490 |
| CTCCAGCAGC | TTCCTAAACC | CTGGGTGGGC | CGTGTTCCAG | CGCTACTGTC | TCACCTGTCC | CACTGTGTCT | 560 |
| TGTCTCAGCG | ACGTAGCTCG | CACGGTTCCT | CCTCACATGG | GGTGTCTGTC | TCCTTCCCCA | ACACTCACAT | 630 |
| GCGTTGAAGG | GAGGAGATTC | TGCGCCTCCC | AGACTGGCTC | CTCTGAGCCT | GAACCTGGCT | CGTGGCCCCC | 700 |
| GATGCAGGTT | CCTGGCGTCC | GGCTGCACGC | TGACCTCCAT | TCCAGGCGC | TCCCCGTCTC | CTGTCACTCTG | 770 |
| CCGGGGCCTG | CCGGTGTGTT | CTTCTGTTTC | TGTGCTCCTT | TCCACGTCCA | GCTGCGTGTG | TCTCTGCCCG | 840 |
| CTAGGGTCTC | GGGGTTTTTA | TAGGCATAGG | ACGGGGGCGT | GGTGGGCCAG | GGCGCTCTTG | GGAAATGCAA | 910 |
| CATTTGGGTG | TGAAAGTAGG | AGTGCCCTGTC | CTCACCTAGG | TCCACGGGCA | CAGGCCTGGG | GATGGAGCCC | 980 |
| CCGCCAGGGA | CCCGCCCTTC | TCTGCCCAGC | ACTTTCCTGC | CCCCCTCCCT | CTGGAACACA | GAGTGGCAGT | 1050 |
| TCCACAAGC | ACTAAGCATC | CTCTTCCCAA | AAGACCCAGC | ATTGGCACCC | CTGGACATTT | GCCCCACAGC | 1120 |
| CCTGGGAATT | CACGTGACTA | CGCACATCAT | GTACACACTC | CCGTCCACGA | CCGACCCCCG | CTGTTTTATT | 1190 |
| TTAATAGCTA | CAAAGCAGGG | AAATCCCTGC | TAAAATGTCC | TTTAAACAAAC | TGGTTAAACA | AACGGGTCCA | 1260 |
| TCCGCACGGT | GGACAGTTCC | TCACAGTGAA | GAGGAACATG | CCGTTTATAA | AGCCTGCAGG | CATCTCAAGG | 1330 |
| GAATTACGCT | GAGTCAAAAC | TGCCACCTCC | ATGGGATACG | TACGCAACAT | GCTCAAAAAG | AAAGAATTTT | 1400 |
| ACCCCATGGC | AGGGGAGTGG | TTAGGGGGGT | TAAGGACGGT | GGGGGCGGCA | GCTGGGGGCT | ACTGCACGCA | 1470 |
| CCTTTTACTA | AAGCCAGTTT | CCTGGTTCTG | ATGGTATTGG | CTCAGTTATG | GGAGACTAAC | CATAGGGGAG | 1540 |
| TGGGGATGGG | GGAACCCGGA | GGCTGTGCCA | TCTTTGCCAT | GCCCGAGTGT | CCTGGGCAGG | ATAATGCTCT | 1610 |
| AGAGATGCCC | ACGTCTTGAT | TCCCCCAAAC | CTGTGGACAG | AACCCGCCCG | GCCCCAGGGC | CTTTCAGGTT | 1680 |
| GTGATCTCCG | TGAGGACCC | GAGGTCTGGG | ATCCTTCGGG | ACTACCTGCA | GGCCGGAAAA | GTAATCCAGG | 1750 |
| GGTTCGTGGA | AGAGGCGGGC | AGGAGGGTCA | GAGGGGGGCA | GCCTCAGGAC | GATGGGAGCA | GTACGTCTGA | 1820 |
| GGCTGAAAA | GGAGGGAGGG | CCTCGAGCCC | AGGCCCTGCA | GCGCCTCCAG | AAGCTGGAAA | AAGCGGGGAA | 1890 |
| GGGACCCCTC | ACGGAGCCCTG | CAGCAGGAAG | GCACGGCTGG | CCCTTAGCCC | ACCAGGGCCC | ATCGTGGACC | 1960 |
| TCCGGCCTCC | GTGCCATAGG | AGGGCACTCG | GCGTGCCCTT | CTAGCATGAA | GTGTGTGGGG | ATTTCAGAA | 2030 |
| GCAACAGGAA | ACCCATGCAC | TGTGAATCTA | GGATTATTTT | AAAACAAAGG | TTTACAGAAA | CATCCAAGGA | 2100 |
| CAGGGCTGAA | GTGCCTCCGG | GCAAGGGCAG | GGCAGGCACG | AGTGATTTTA | TTTAGCTATT | TTATTTTATT | 2170 |
| TACTTACTTT | CTGAGACAGA | GTTATGCTCT | TGTTGCCCAG | GCTGGAGTGC | AGCGGCATGA | TCTTGGCTCA | 2240 |
| CTGCAACCTC | CGTCTCCTGG | GTTCAAGCAA | TTCTCGTGCC | TCAGCCTCCC | AAGTAGCTGG | GATTTCAGGC | 2310 |
| GTGCACCACC | ACACCCGGCT | AATTTTGTAT | TTTTAGTAGA | GATGGGCTTT | CACCATGTTG | GTCAAGCTGA | 2380 |
| TCTCAAAATC | CTGACCTCAG | GTGATCCGCC | CACCTCAGCC | TCCCAAAGTG | CTGGGATTAC | AGGCATGAGC | 2450 |
| CACTGCACCT | GGCCTATTTA | ACCATTTTAA | AACCTCCCTG | GGCTCAAGTC | ACACCCACTG | GTAAGGAGTT | 2520 |
| CATGGAGTTC | AATTTCCCCT | TTACTCAGGA | GTTACCCTCC | TTTGATATTT | TCTGTAATTC | TTCTGAGACT | 2590 |
| GGGGATACAC | CGTCTCTTGA | CATATTCACA | GTTTCTGTGA | CCACCTGTTA | TCCCATGGGA | CCCACCTCAG | 2660 |
| GGGCAGCTGG | GAGGCTGCAG | GCTTCAGGTC | CCAGTGGGGT | TGCCATCTGC | CAGTAGAAAC | CTGATGTAGA | 2730 |
| ATCAGGGCGC | AAGTGTGGAC | ACTGTCCTGA | ATCTCAATGT | CTCAGTGTGT | GCTGAAACAT | GTAGAAATTA | 2800 |
| AAGTCCATCC | CTCCTACTCT | ACTGGGATTG | AGCCCCTTCC | CTATCCCCCC | CCAGGGGCAG | AGGAGTTCCT | 2870 |
| CTCACTCCTG | TGGAGGAAGG | AATGATACTT | TGTTATTTTT | CACTGCTGGT | ACTGAATCCA | CTGTTTCATT | 2940 |
| TGTTGGTTTG | TTTGTTTTGT | TTTGAGAGGC | GGTTTCACTC | TTGTTGCTCA | GGCTGGAGGG | AGTGCAATGG | 3010 |
| CGCGATCTTG | GCTTACTGCA | GCCTCTGCCT | CCCAGGTTCA | AGTGATTCTC | CTGCTTCCGC | CTCCCATTTG | 3080 |
| GCTGGGATTA | CAGGCACCCG | CCACCATGCC | CAGCTAATTT | TTTGATTTTT | TAGTAGAGAC | GGGGGTGGGT | 3150 |
| GGGGTTCAAC | ATGTTGGCCA | GGCTGGTCTC | GAACCTTCTGA | CCTCAGATGA | TCCACCTGCC | TCTGCCTCCT | 3220 |
| AAAGTGCTGG | GATTACAGGT | GTGAGCCACC | ATGCCCAGCT | CAGAATTTAC | TCTGTTTAGA | AACATCTGGG | 3290 |
| TCTGAGGTAG | GAAGCTCACC | CCACTCAAGT | GTTGTGGTGT | TTTAAGCCAA | TGATAGAATT | TTTTTATTGT | 3360 |
| TGTTAGAAC | CTCTTGATGT | TTTACACTGT | GATGACTAAG | ACATCATCAG | CTTTTCAAAG | ACACACTAAC | 3430 |
| TGCACCCATA | ATACTGGGGT | GTCTTCTGGG | TATCAGCAAT | CTTCATTGAA | | | |

Fig. 4 (Fortsetzung)

CACAGCCTAG GCCGATTGGA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA 4760
GCGGCGCGCG GCGGGGAAG CGCGGCCAG ACCCCCGGGT CCGCCCGGAG CAGCTGCGCT GTCGGGGCCA 4830
GGCCGGGCTC CCA GTGGATT CGCGGGCACA GACGCCCAGG ACCGCGCTCC CCACGTGGCG GAGGGACTGG 4900
GGACCCGGGC ACCCGTCCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCCC GCGCGTCCC 4970
GACCCCTCCC GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCCT CCCCTTCCTT TCCGCGGCCC 5040
CGCCCTCTCC TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCGC ACGTGGGAAG CCCTGGCCCC 5110
GGCCACCCCC GCGATG 5126

6 / 15

Fig. 5

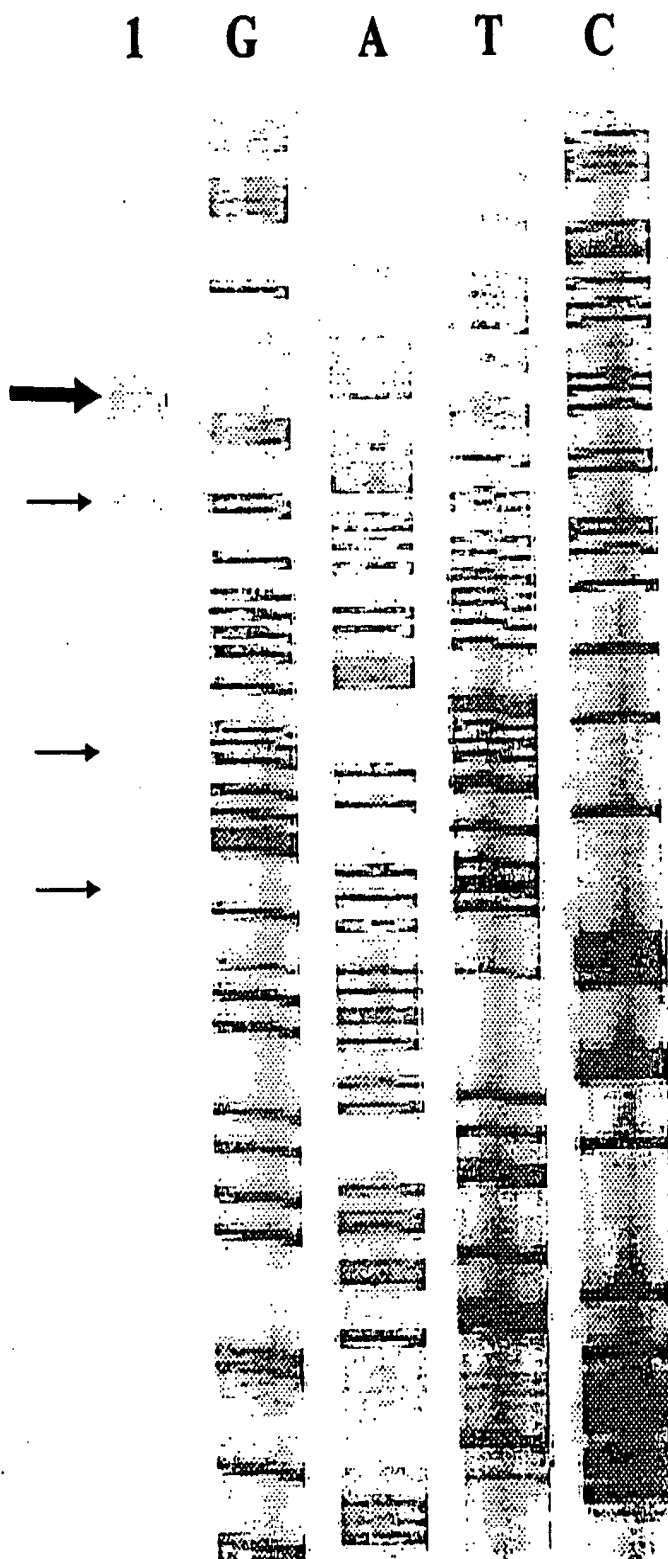


Fig. 6

| | | | | | | | |
|-------------|-------------|------------|------------|-------------|------------|------------|------|
| GTTCAGGCA | GCGCTGCGTC | CTGCTGCGCA | CGTGGGAAGC | CCTGGCCCCG | GCCACCCCCG | CGATGCCGCG | 70 |
| CGCTCCCCGC | TGCCGAGCCG | TGCGCTCCCT | GCTGCGCAGC | CACTACCGCG | AGGTGCTGCC | GCTGGCCACG | 140 |
| TTCTGTGCGC | GCCTGGGGCC | CCAGGGCTGG | CGGCTGGTGC | AGCGCGGGGA | CCCGGCGGCT | TTCGCGCGC | 210 |
| TGGTGGCCCA | GTGCTGGTG | TGCGTGCCCT | GGGACGCACG | GCCGCCCCC | GCCGCCCCCT | CCTTCCGCCA | 280 |
| GGTGTCTTGC | CTGAAGGAGC | TGGTGGCCCC | AGTGCTGCAG | AGGCTGTGCG | AGCGCGGCGC | GAAGAAGCTG | 350 |
| CTGGCCCTTCG | GCTTCGCGCT | GCTGGACGGG | GCCCCGCGGG | GCCCCCCCCA | GGCCTTCACC | ACCAGCGTGC | 420 |
| GCAGCTACCT | GCCCCAACACG | GTGACCGACG | CACTGCGGGG | GAGCGGGGCG | TGGGGGCTGC | TGCTGCGCCG | 490 |
| CGTGGGCGAC | GACGTGCTGG | TTCACCTGCT | GGCACGCTGC | GCGCTCTTTG | TGCTGGTGCG | TCCCAGCTGC | 560 |
| GCCTACCAGG | TGTGCGGGCC | GCCGCTGTAC | CAGCTCGGCG | CTGCCACTCA | GGCCCCGGCC | CCGCCACACG | 630 |
| CTAGTGGACC | CCGAAGGCGT | CTGGGATGCG | AACGGGCCTG | GAACCATAGC | GTCAGGGAGG | CCGGGGTCCC | 700 |
| CCTGGGCTTG | CCAGCCCCCG | GTGCGAGGAG | GCGCGGGGGC | AGTGCCAGCC | GAAGTCTGCC | GTTGCCCAAG | 770 |
| AGGCCCAGGC | GTGGCGCTGC | CCCTGAGCCG | GAGCGGACGC | CCGTTGGGCA | GGGGTCTCTG | GCCCACCCGG | 840 |
| GCAGGACGCG | TGGACCGAGT | GACCGTGGTT | TCTGTGTGGT | GTCACCTGCC | AGACCCGCGG | AAGAAGCCAC | 910 |
| CTCTTTGGAG | GGTGCCTCT | CTGGCACGCG | CCACTCCCAC | CCATCCGTGG | GCCGCCAGCA | CCACGCGGGC | 980 |
| CCCCATCCA | CATCGCGGCC | ACCACGTCCC | TGGGACACGC | CTTGTCCCCC | GGTGTACGCC | GAGACCAAGC | 1050 |
| ACTTCCTCTA | CTCCTCAGGC | GACAAGGAGC | AGCTGCGGCC | CTCCTTCCTA | CTCAGCTCTC | TGAGGCCCAG | 1120 |
| CCTGACTGGC | GCTCGGAGGC | TCGTGGAGAC | CATCTTTCTG | GGTTCCAGGC | CCTGGATGCC | AGGGACTCCC | 1190 |
| CGCAGGTTGC | CCCGCCTGCC | CCAGCGCTAC | TGGCAAATGC | GGCCCCGTGT | TCTGGAGCTG | CTTGGGAACC | 1260 |
| ACGCGCAGTG | CCCCTACGGG | GTGCTCCTCA | AGACGCACTG | CCCGCTGCGA | GCTGCGGTCA | CCCCAGCAGC | 1330 |
| CGGTGTCTGT | GCCCCGGGAGA | AGCCCCAGGG | CTCTGTGGCG | GCCCCCGAGG | AGGAGGACAC | AGACCCCCGT | 1400 |
| CGCCTGGTGC | AGCTGCTCCG | CCAGCACAGC | AGCCCCTGGC | AGGTGTACGG | CTTCGTGCGG | GCCTGCCTGC | 1470 |
| GCCGGCTGGT | GCCCCCAGGC | CTCTGGGGCT | CCAGGCACAA | CGAACGCCGC | TTCCTCAGGA | ACACCAAGAA | 1540 |
| GTTTCATCTC | CTGGGGAAGC | ATGCCAAGCT | CTCGTGCAG | GAGCTGACGT | GGAAGATGAG | CGTGCGGGAC | 1610 |
| TGCGCTTGGC | TGCGCAGGAG | CCCAGGGGTT | GGCTGTGTTT | CGGCCGAGA | GCACCGTCTG | CGTGAGGAGA | 1680 |
| TCTTGCCAA | GTTCTGTCAC | TGGCTGATGA | GTGTGTACGT | CGTCGAGCTG | CTCAGGTCTT | TCTTTTATGT | 1750 |
| CACGGAGACC | ACGTTTCAA | AGAACAGGCT | CTTTTTCTAC | CGGAAGAGTG | TCTGGAGCAA | GTTGCAAAGC | 1820 |
| ATTGGAATCA | GACAGCACTT | GAAGAGGGTG | CAGCTGCGGG | AGCTGTGCGA | AGCAGAGGTC | AGGCAGCATC | 1890 |
| GGGAAGCCAG | GCCCGCCCTG | CTGACGTCCA | GACTCCGCTT | CATCCCCAAG | CCTGACGGGC | TGCGGCCGAT | 1960 |
| TGTGAACATG | GACTACGTCG | TGGGAGCCAG | AACGTTCCGC | AGAGAAAAGA | GGGCCGAGCG | TCTCACCTCG | 2030 |
| AGGGTGAAGG | CACGTGTCAG | CGTGCTCAAC | TACGAGCGGG | CGCGGCGCCC | CGGCCTCTG | GGCGCCTCTG | 2100 |
| TGCTGGGCCT | GGACGATATC | CACAGGGCCT | GGCGCACCTT | CGTGCTGCGT | GTGCGGGCCC | AGGACCCGCC | 2170 |
| GCCTGAGCTG | TACTTTGTCA | AGGTGGATGT | GACGGGCGCG | TACGACACCA | TCCCCCAGGA | CAGGCTCAGC | 2240 |
| GAGGTCATCG | CCAGCATCAT | CAAACCCAG | AACACGTACT | GCGTGCGTCG | GTATGCCGTG | GTCCAGAAGG | 2310 |
| CCGCCCATGG | GCACGTCCGC | AAGGCCTTCA | AGAGCCACGT | CTCTACCTTG | ACAGACCTCC | AGCCGTACAT | 2380 |
| GCGACAGTTC | GTGGCTCACC | TGCAGGAGAC | CAGCCCGCTG | AGGGATGCCG | TCGTATCGA | GCAGAGCTCC | 2450 |
| TCCCTGAATG | AGGCCAGCAG | TGGCCTCTTC | GACGTCTTCC | TACGCTTCAT | GTGCCACCAC | GCCGTGCGCA | 2520 |
| TCAGGGGCAA | GTCTACGTC | CAGTGCCAGG | GGATCCCGCA | GGGTCCATC | CTCTCCACGC | TGCTCTGCAG | 2590 |
| CCTGTGCTAC | GGCGACATGG | AGAACAAGCT | GTTTGCGGGG | ATTGCGCGGG | ACGGGCTGCT | CCTGCGTTTG | 2660 |
| GTGGATGATT | TCTTGTGGT | GACACCTCAC | CTCACCCACG | CGAAACCTT | CCTCAGGACC | CTGGTCCGAG | 2730 |
| GTGTCCCTGA | GTATGGCTGC | GTGGTGAAGT | TGCGGAAGAC | AGTGGTGAAC | TTCCCTGTAG | AAGACGAGGC | 2800 |
| CCTGGGTGGC | ACGGCTTTTG | TTCAGATGCC | GGCCACGGCG | CTATTCCCCT | GGTGCGGCCT | GCTGCTGGAT | 2870 |
| ACCCGGACCC | TGGAGGTGCA | GAGCGACTAC | TCCAGCTATG | CCCGGACCTC | CATCAGAGCC | AGTCTCACCT | 2940 |
| TCAACCGCGG | CTTCAAGGCT | GGGAGGAACA | TGCGTCGCAA | ACTCTTTGGG | GTCTTGCGGC | TGAAGTGTCA | 3010 |
| CAGCCTGTTT | CTGGATTTCG | AGGTGAACAG | CCTCCAGACG | GTGTGCACCA | ACATCTACAA | GATCCTCTCG | 3080 |
| CTGCAGGCGT | ACAGGTTTCA | CGCATGTGTG | CTGCAGCTCC | CATTTTCATCA | GCAAGTTTGG | AAGAACCCCA | 3150 |
| CATTTTTCCT | GCGCGTCATC | TCTGACACGG | CCTCCCTCTG | CTACTCCATC | CTGAAAGCCA | AGAACGCAGG | 3220 |
| GATGTCGCTG | GGGGCCAAGG | GCGCGCGCGG | CCCTCTGCCC | TCCGAGGCGG | TGCAGTGGCT | GTGCCACCAA | 3290 |
| GCATTCTGTC | TCAAGCTGAC | TCGACACCGT | GTCACCTACG | TGCCACTCCT | GGGGTCACTC | AGGACAGCCC | 3360 |
| AGACGCAGCT | GAGTCGGAAG | CTCCCGGGGA | CGACGCTGAC | TGCCCTGGAG | GCCGCAGCCA | ACCCGGCACT | 3430 |
| GCCCTCAGAC | TTCAAGACCA | TCCTGGACTG | ATGGCCACCC | GCCACAGCC | AGGCCGAGAG | CAGACACCAG | 3500 |
| CAGCCCTGTC | ACGCCGGGCT | CTACGTCCCA | GGGAGGGAGG | GGCGGCCCCAC | ACCCAGGCCC | GCACCGCTGG | 3570 |
| GAGTCTGAGG | CCTGAGTGAG | TGTTTGGCCG | AGGCCTGCAT | GTCCGGCTGA | AGGCTGAGTG | TCCGGCTGAG | 3640 |
| GCCTGAGCGA | GTGTCCAGCC | AAGGGCTGAG | TGTCCAGCAC | ACCTGCCGTC | TTCATTCCC | CACAGGCTGG | 3710 |
| CGCTCGGCTC | CACCCAGGG | CCAGCTTTTC | CTCACCAGGA | GCCCGGCTTC | CACTCCCCAC | ATAGGAATAG | 3780 |

Fig. 7

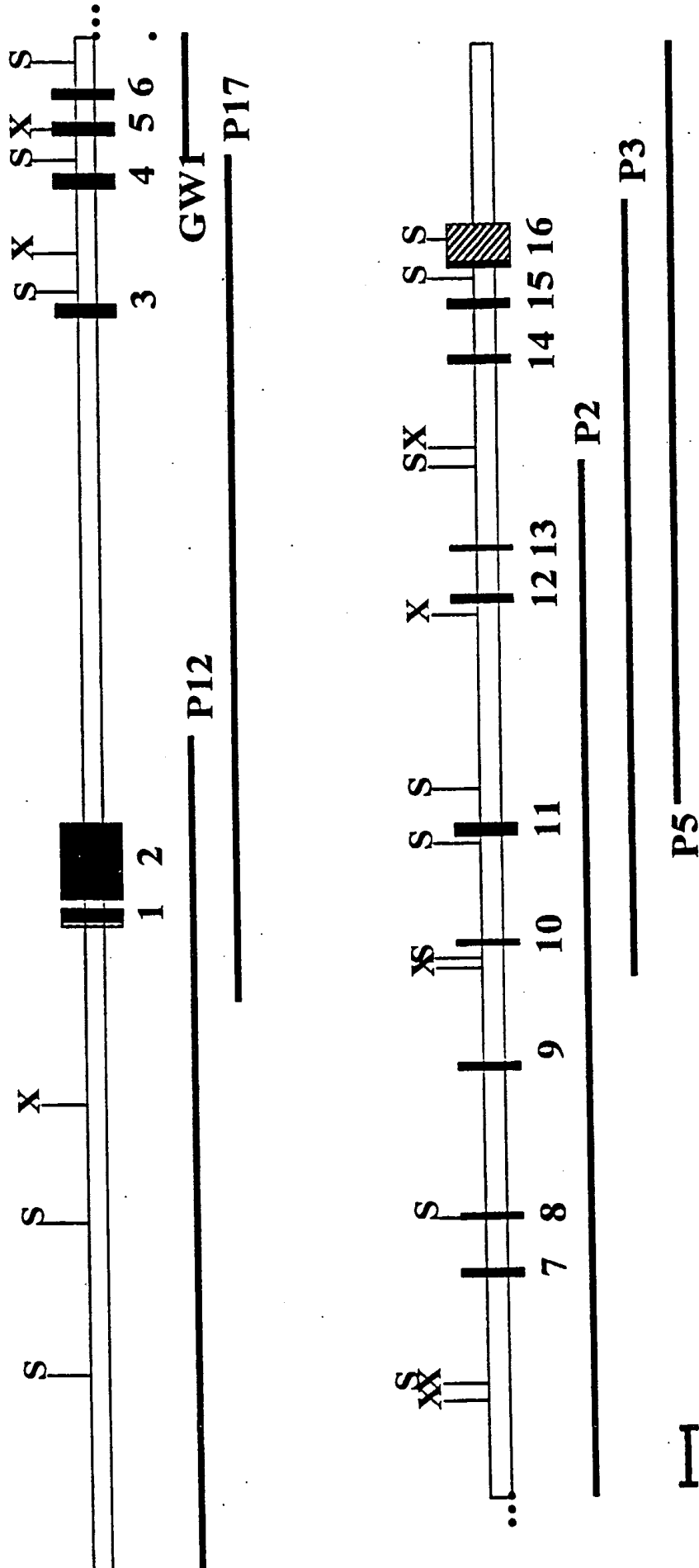


Fig. 8A

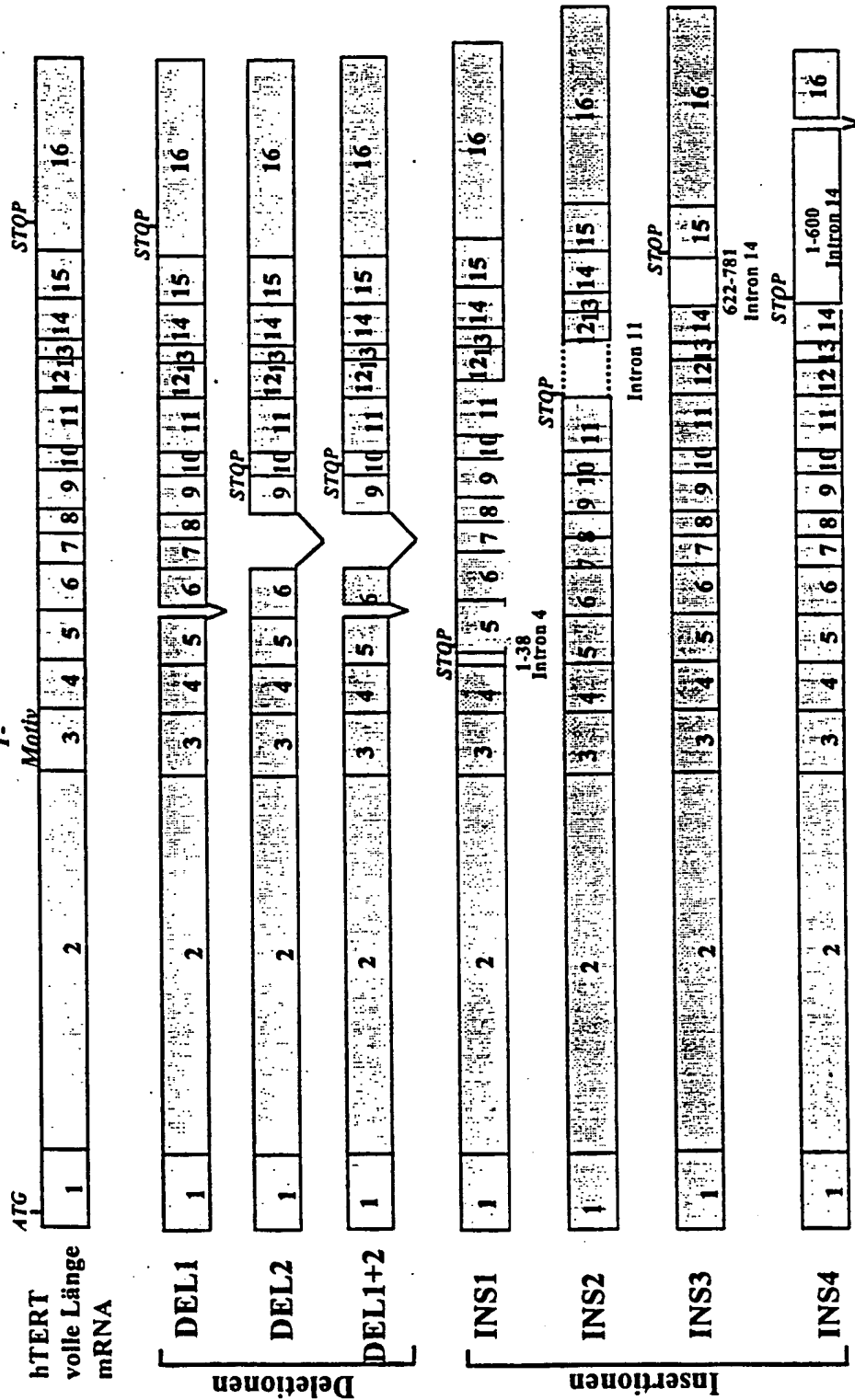


Fig. 8B

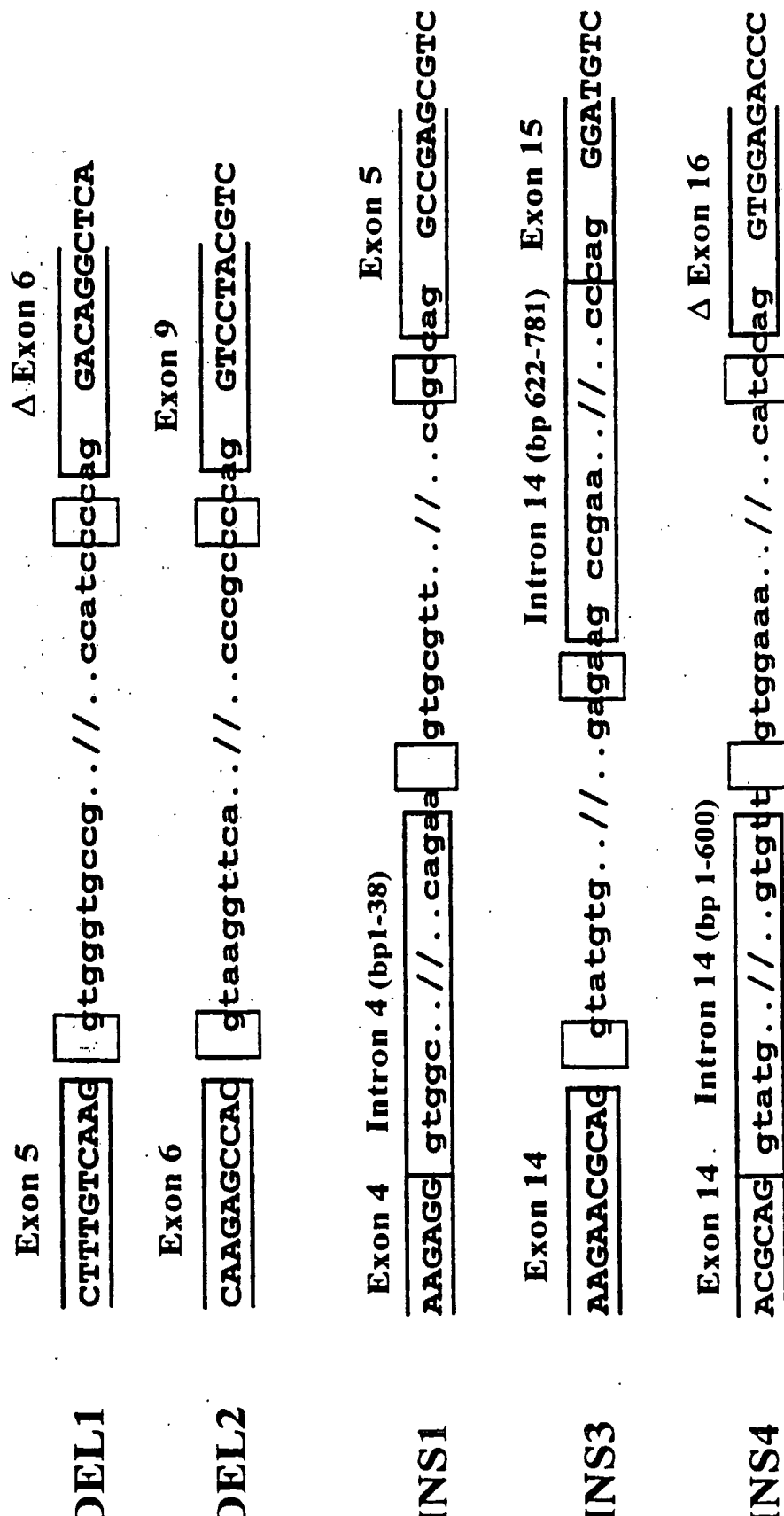


Fig. 9

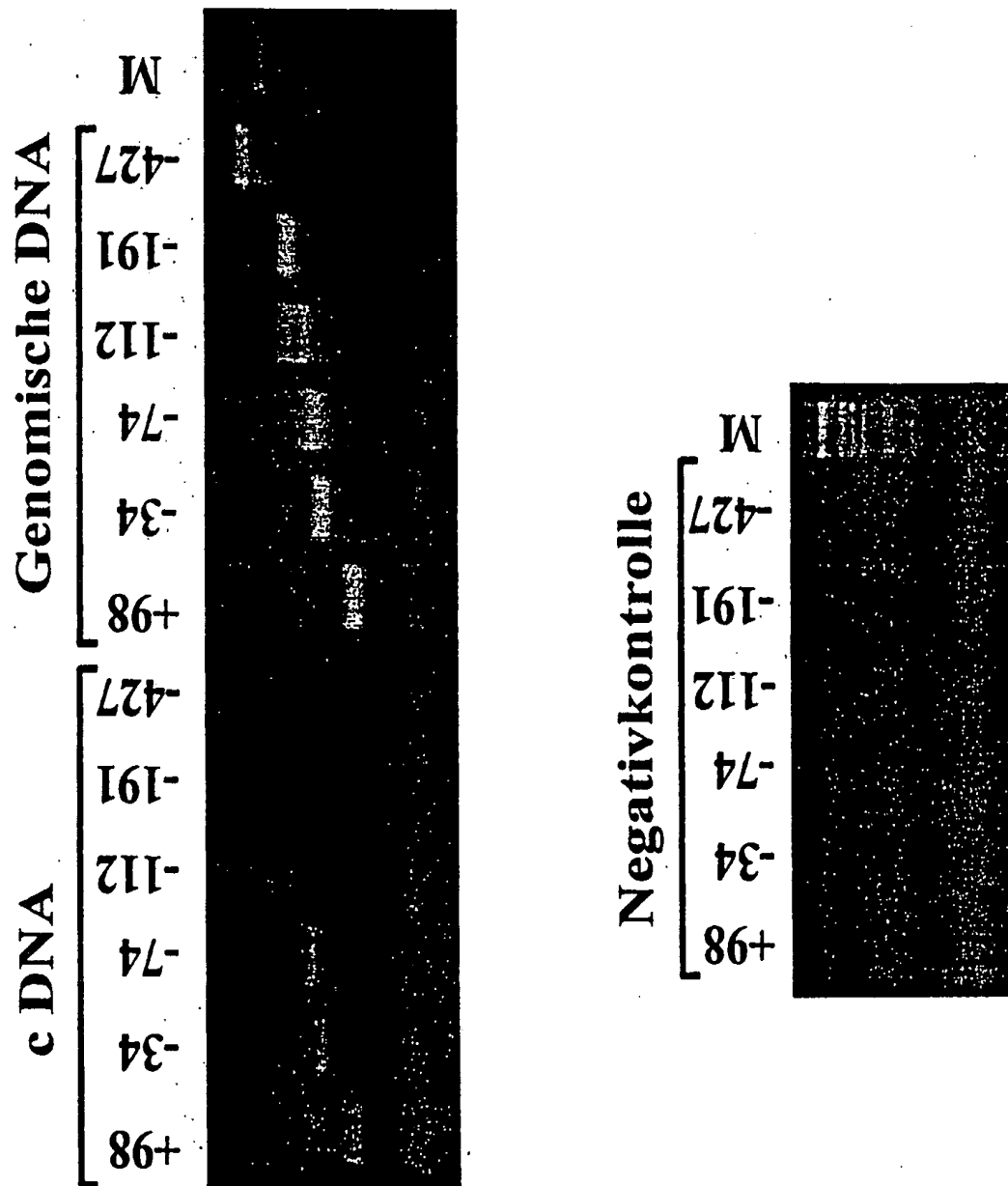


Fig. 10

| | | | | | | | |
|-------------|-------------|------------|-------------|-------------|-------------|-------------|--------|
| ACTTGAGCCC | AAGAGTTCAA | GGCTACGGTG | AGCCATGATT | GCAACACCAC | ACGCCAGCCT | TGGTGACAGA | -11204 |
| ATGAGACCCT | GTCTCAAAAA | AAAAAAGGAA | AATTGAAATA | ATATAAAGCA | TCTTCTCTGG | CCACAGTGGA | -11134 |
| ACAAAACCAG | AAATCAACAA | CAAGAGGAAT | TTTGAAAAC | ATACAAACAC | ATGAAAAATTA | AACAATATAC | -11064 |
| TTCTGAATGA | CCAGTGAGTC | AATGAAGAAA | TTAAAAAGGA | AATTGAAAAA | TTTATTTAAG | CAATGATATA | -10994 |
| CGGAAACATA | ACCTCTCAAA | ACCCACGGTA | TACAGCAAAA | GCAGTGCTAA | GAAGGAAGTT | TATAGCTATA | -10924 |
| AGCAGCTACA | TCAAAAAAGT | AGAAAAGCCA | GGCGCAGTGG | CTCATGCCTG | TAATCCCAGC | ACTTTGGGAG | -10854 |
| GCCAAGGCGG | GCAGATCGCC | TGAGGTGAGG | AGTTTCGAGC | CAGCCTGACC | AACACAGAGA | AACCTTGTCG | -10784 |
| CTACTAAAAA | TACAAAAATTA | GCTGGGCATG | GTGGGCACATG | CCTGTAATCC | CAGCTACTCG | GGAGGCTGAG | -10714 |
| GCAGGATAAC | CGCTTGAACC | CAGGAGGTGG | AGGTTGCGGT | GAGCCGGGAT | TGCGCCATTG | GACTCCAGCC | -10644 |
| TGGGTAACAA | GAGTGAAACC | CTGTCTCAAG | AAAAAAGAAA | AAGTAGAAAA | ACTTAAAAAT | ACAACCTAAT | -10574 |
| GATGCACCTT | AAAGAACTAG | AAAAGCAAGA | GCAAACTAAA | CCTAAAATTG | GTAAAAGAAA | AGAAATAATA | -10504 |
| AAGATCAGAG | CAGAAATAAA | TGAAACTGAA | AGATAACAAT | ACAAAAGATC | AACAAAATTA | AAAGTTGGTT | -10434 |
| TTTTGAAAAG | ATAAACAAAA | TTGACAAACC | TTTGCCCAAG | CTAAGAAAAA | AGGAAAGAAG | ACCTAAATAA | -10364 |
| ATAAAGTCAG | AGATGAAAAA | AGAGACATTA | CAACTGATAC | CACAGAAATT | CAAAGGATCA | CTAGAGGCTA | -10294 |
| CTATGAGCAA | CTGTACACTA | ATAAATTGAA | AAACTGTAGAA | AAATCTAGATA | AATCCCTAGA | TGCATACAAAC | -10224 |
| CTACCAAGAT | TGAACCATGA | AGAAATCCAA | AGCCCAAAACA | GACCAATAAC | AATAATGGGA | TTAAAGCCAT | -10154 |
| AATAAAAAAGT | CTCCTAGCAA | AGAGAAGCCC | AGGACCCAAT | GGCTTCCCTG | CTGGATTTTA | CCAATCATTT | -10084 |
| AAAGAAGAAT | GAATTCCAAT | CCTACTCAAA | CTATTCTGAA | AAATAGAGGA | AAGAATACTT | CCAACTCAT | -10014 |
| TCTACATGGC | CAGTATTACC | CTGATTCCAA | AACCAGACAA | AAACACATCA | AAAACAAACA | AACAAAAAAA | -9944 |
| CAGAAAGAAA | GAAGAACTACA | GGCCAATATC | CCTGATGAAT | ACTGATACAA | AAATCCTCAA | CAAAACACTA | -9874 |
| GCAAAACCAA | TTAAACAACA | CCTTCGAAAG | ATCATTCATT | GTGATCAAGT | GGGATTTTAT | CCAGGGATGG | -9804 |
| AAGGATGGTT | CAACATATGC | AAATCAATCA | ATGTGATACA | TCATCCCAAC | AAAATGAAGT | ACAAAACTA | -9734 |
| TATGATTATT | TCACCTTTATG | CAGAAAAAGC | ATTTGATAAA | ATTCTGCACC | CTTCATGATA | AAAACCTCA | -9664 |
| AAAAACCAGG | TATACAAGAA | ACATACAGGC | CAGGCACAGT | GGTCCACACC | TGCGATCCCA | GCACTCTGGG | -9594 |
| AGGCCAAGGT | GGGATGATTG | CTTGGGCCCC | GGAGTTTGAG | ACTAGCCTGG | GCAACAAAAT | GAGACCTGGT | -9524 |
| CTACAAAAAA | CTTTTTTTAAA | AAATTAGCCA | GGCATGATGG | CATATGCCTG | TAGTCCCAGC | TAGTCTGGAG | -9454 |
| GCTGAGGTGG | GAGAATCACT | TAAGCCTAGG | AGGTCCAGGC | TGCAGTGAGC | CATGAACATG | TCAGTGTACT | -9384 |
| CCAGCCTAGA | CAACAGAACA | AGACCCCACT | GAATAAGAAG | AAGGAGAAGG | AGAAGGGAGA | AGGGAGGGAG | -9314 |
| AAGGGAGGAG | GAGGAGAAGG | AGGAGGTGGA | GGAGAAGTGG | AAGGGGAAGG | GGAAGGGAAA | GAGGAAGAAG | -9244 |
| AAGAAACATA | TTTCAACATA | ATAAAGCCCC | TATATGACAG | ACCGAGGTAG | TATTATGAGG | AAAAACTGAA | -9174 |
| AGCCTTTCCCT | CTAAGATCTG | GAAAAAGACA | AGGGCCCACT | TTCAACCACTG | TGATTCAACA | TAGTACTAGA | -9104 |
| AGTCTTAGCT | AGAGCAATCA | GATAAGAGAA | AGAAATAAAA | GGCATCCAAA | CTGGAAAGGA | AGAAGTCAAA | -9034 |
| TTATCCTGTT | TGCAGATGAT | ATGATCTTAT | ATCTGGAATA | GACTTAAGAC | ACCATAAAA | AACTATTAGA | -8964 |
| GCTGAAATTT | GGTACAGCAG | GATACAAAAT | CAATGTACAA | AAATCAGTAG | TATTTCTATA | TTCCAACAGC | -8894 |
| AAACAATCTG | AAAAAGAAAC | CAAAAAAGCA | GCTACAAATA | AAATTAACA | GCTAGGAATT | AACCAAGAA | -8824 |
| GTGAAAGATC | TCTACAATGA | AACTATAAAA | ATGTTGATAA | AAGAAATTGA | AGAGGGCACA | AAAAAAGAAA | -8754 |
| AGATATTCCA | TGTTCATAGA | TTGGAAGAAT | AAATACTGTT | AAAATGTCCA | TACTACCCAA | AGCAATTTAC | -8684 |
| AAATTCATG | CAATCCCTAT | TAAAATACTA | ATGACGTTCT | TCACAGAAAT | AGAAGAAACA | ATTCTAAGAT | -8614 |
| TTGTACAGAA | CCACAAAAGA | CCGAGAATAG | CTTGACCAAT | CTGACCAAAA | AAGAACAACA | CTGGAAGCAT | -8544 |
| CACATTACCT | GACTTCAAT | TATACTACAA | AGCTATAGTA | ACCCAAACTA | CATGGTACTG | GCATAAAAC | -8474 |
| AGATGAGACA | TGGACCAGAG | GAACAGAATA | GAGAATCCAG | AAACAAATCC | ATGCATCTAC | AGTGAACCTA | -8404 |
| TTTTTGACAA | AGGTGCCAAG | AACATACTTT | GGGGAAGAGA | TAATCTCTTC | AATAAATGGT | GCTGGAGGAA | -8334 |
| CTGGATATCC | ATATGCAAAA | TAACATACT | AGAATCTGT | CTCTCACCAT | ATACAAAAGC | AAATCAAAAT | -8264 |
| GGATGAAAGG | CTTAAATCTA | AAACCTCAAA | CTTTGCAACT | ACTAAAAGAA | AACACCGGAG | AAACTCTCCA | -8194 |
| GGACATTGGA | TGTTGCAAGG | ACTTCTTGAG | TAATCTCCTG | CAGGCACAGG | CAACCAAAGC | AAAAACAGAC | -8124 |
| AAATGGGATC | ATATCAAGTT | AAAAAGCTTC | TGCCCAAGCA | AGGAAACAAT | CAACAAAGAG | AAGAGACAAC | -8054 |
| CCACAGAATG | GGAGAATATA | TTTGCAAACT | ATTCTATCTA | CAAGGAATTA | ATAACCAGTA | TATATAAGGA | -7984 |
| GCTCAAACTA | CTCTATAAGA | AAAACACCTA | ATAAGCTGAT | TTTCAAAAAT | AAGCAAAAGA | TCTGGGTAGA | -7914 |
| CATTTCTCAA | AATAAGTCAT | ACAAATGGCA | AACAGGCATC | TGAAAATGTG | CTCAACACCA | CTGATCATCA | -7844 |
| GAGAAATGCA | AATCAAAACT | ACTATGAGAG | ATCATCTCAT | CCCAGTTAAA | ATGGCTTTTA | TTCAAAAGAC | -7774 |
| AGGCAATAAC | AAATGCCAGT | GAGGATGTGG | ATAAAGGAA | ACCTTGGAC | ACTGTTGGTG | GGATGGAAAT | -7704 |
| TTGCTACCAC | TATGGAGAAC | AGTTTGAAAG | TTCTCAAAA | AACCTAAAAAT | AAAGCTACCA | TACAGCAATC | -7634 |
| CCATTGCTAG | GTATATACTC | CAAAAAAGGG | AATCAGTGTA | TCAACAAGCT | ATCTCCACTC | CCACATTTAC | -7564 |
| TGCAGCACTG | TTCATAGCAG | CCAAGGTTTG | GAAGCAACCT | CAGTGCCAT | CAACAGACGA | ATGGAAAAAG | -7494 |
| AAAATGTGGT | GCACATACAC | AATGGAGTAC | TACGCAGCCA | TAAAAAAGAA | TGAGATCCTG | TCAGTTGCAA | -7424 |
| CAGCATGGGG | GGCACTGGTC | AGTATGTTAA | GTGAAATAAG | CCAGGCACAG | AAAGACAAAC | TTTTCATGTT | -7354 |
| CTCCCTTACT | TGTGGGAGCA | AAAATTAATA | CAATTGACAT | AGAAATAGAG | GAGAATGGTG | GTTCTAGAGG | -7284 |
| GGTGGGGGAC | AGGGTGACTA | GAGTCAACAA | TAAATTTATTG | TATGTTTTAA | AATAACTAAA | AGAGTATAAT | -7214 |
| TGGGTTGTTT | GTAACACAAA | GAAAGGATAA | ATGCTTGAAG | GTGACAGATA | CCCCATTTAC | CCTGATGTGA | -7144 |
| TTATTACACA | TTGTATGCCT | GTATCAAAAT | ATCTCATGTA | TGCTATAGAT | ATAAACCTTA | CTATATTTAA | -7074 |
| AATTAATAAT | TTAATGGCCA | GGCACGGTGG | CTCATGTCCG | TAATCCCAGC | ACTTTGGGAG | GGCGAGGCGG | -7004 |
| GTGGATCACC | TGAGGTCAGG | AGTTTGAAAC | CAGTCTGGCC | ACCATGATGA | AACCCTGTCT | CTACTAAAGA | -6934 |
| TACAAAAAAT | TAGCCAGGCGT | GGTGGCACAT | ACCTGTAGTC | CCAATCTACT | AGGAGGCTGA | GACAGGAGAA | -6864 |
| TTGCTTGAAC | CTGGGAGGCG | GAGGTTGCAG | TGAGCCGAGA | TGATGCCACT | GCACTGCAGC | CTGGGTGACA | -6794 |
| GAGCAAGACT | CCATCTCAAA | ACAAAAACAA | AAAAAAGAA | ATTAAAAATTG | TAATTTTTAT | GTACCGTATA | -6724 |
| AATATATACT | CTACTATATT | AGAAGTTAAA | AATTAACA | ATTATAAAG | GTAATTAACC | ACTTAATCTA | -6654 |
| AAATAAGAAC | AATGTATGTG | GGGTTTCTAG | CTTCTGAAGA | AGTAAAGTT | ATGGCCACGA | TGGCAGAAAT | -6584 |

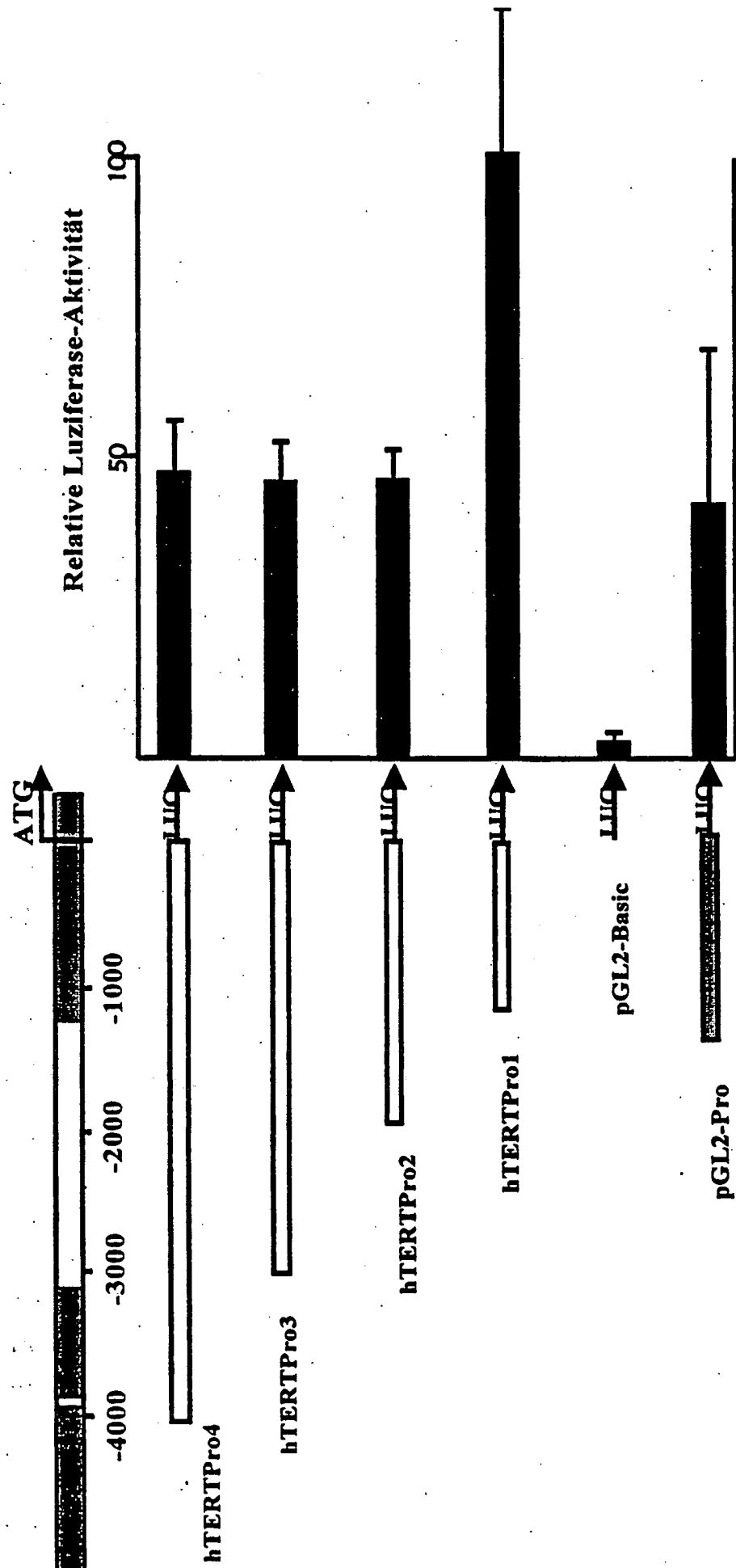
Fig. 10

GTGAGGAGGG AACAGTGGA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTTAACC -6514
 AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAACAA CTGCTAATAA TGGTGAAAGG -6444
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA -6374
 CACCGTCCTC TCATTACCGG TGCTTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGTG TTCGTGTTTG -6304
 GTTAAACTTA ATCTGTATGA ATCCTGAAAC GAAAAATGGT GGTGATTTC TCCAGAAGAA TTAGAGTACC -6234
 TGGCAGGAAG CAGGTGGCTC TGTGGACCTG AGCCACTTCA ATCTTCAAGG GTCTCTGGCC AAGACCCAGG -6164
 TGCAAGGCAG AGGCCTGATG ACCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT -6094
 CGTTGGTGAG CAGCGCATGA AGTGCCCTTA TTTACGCTTT GCAAAGATTG CTCTGGATAC CATCTGAAA -6024
 AGGCGGCCAG CGGGAATGCA AGGAGTCAGA AGCCTCCTGC TCAAACCCAG GCCAGCAGCT ATGGCGCCCA -5954
 CCCGGGCGTG TGCCAGAGGG AGAGGAGTCA AGGCACCTCG AAGTATGGCT TAAATCTTTT TTTACCTGA -5884
 AGCAGTGACC AAGGTGTATT CTGAGGGAAG CTTGAGTTAG GTGCCTTCTT TAAAACAGAA AGTCATGGAA -5814
 GCACCTTCT CAAGGGA AAAACAGGCCC GCTCTGCGGT CATTTACCTC TTTCTCTCT CCCTCTCTTG -5744
 CCCTCGCGGT TTCTGATCGG GACAGAGTGA CCCCCGTGGA GCTTCTCCGA GCGGCTGTG AGGACCTCT -5674
 TGCAAGAGGC TCCACAGACC CCCGCCCTGG AGAGAGGAGT CTGAGCCTGG CTTAATAACA AACTGGGATG -5604
 TGGCTGGGG CGGACAGCGA CGGCGGGATT CAAAGACTTA ATTCCATGAG TAAATTC AAC CTTTCCACAT -5534
 CCGAATGGAT TTGGATTTTA TCTTAATATT TTCTTAAATT TCATCAAATA ACATTACAGG CTGCAGAAAT -5464
 CCAAAGGCGT AAAACAGGAA CTGAGCTATG TTTGCCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG -5394
 ATTTTTCGCG CTAAGTACTT TTTATTGGTT TTCATAAGGT GGCTTAGGGT GCAAGGGAAG GTACACGAGG -5324
 AGAGGCGTGG CGGCGAGGGC TATGAGCAG GCAGGGCCAC CGGGGAGAGA GTCCCCGGCC TGGGAGGCTG -5254
 ACAGCAGGAC CACTGACCGT CCTCCCTGGG AGCTGCCACA TTGGGCAACG CGAAGGCGGC CACGCTGCGT -5184
 GTGACTCAGG ACCCATACC GGCTTCCTGG GCCCACCAC ACTAACCCAG GAAGTCACGG AGCTCTGAAC -5114
 CCGTGGAAC GAACATGACC CTGCGCTGCC TGCTTCCTG GGTGGGTCAA GGGTAATGAA GTGGTGTCGA -5044
 GGAAATGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATTATT CATCTTCAAC -4974
 CCCAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGA CAAGCCATGA CAAAACCTAG TACAAACACC -4904
 ACTCTTTTAC TAGGCCACA GAGCACGGSC CACACCCCTG ATATATTAAG AGTCCAGGAG AGATGAGGCT -4834
 GCTTTACGCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTTCC TCTAGACTAG TAGACCCTGG -4764
 CAGGCACTCC CCCAGATTCT AGGGCCTGGT TGCTGCTTCC CGAGGGCGCC ATCTGCCCTG GAGACTCAGC -4694
 CTGGGGTGCC AACTGAGGC CAGCCCTGTC TCCACACCCT CCGCTCCAG GCCTCAGCTT CTCGAGCAGC -4624
 TTCTTAAACC CTGGGTGGGC CGTGTTCCAG CGCTACTGTC TCACCTGTCT CACTGTGTCT TGCTCAGCG -4554
 ACGTAGCTCG CACGGTTCTC CTCACATGG GGTGTCTGTC TCCTTCCCA AACTCACAT GCGTTGAAGG -4484
 GAGGAGATT TGCGCCTCCC AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCCC GATGAGGTT -4414
 CCTGGCGTCC GGCTGCACGC TGACCTCCAT TTCCAGGCGC TCCCCGTCTC CTGTCATCTG CCGGGGCTG -4344
 CCGGTGTGTT CTTCTGTTTC TGTGCTCCTT TCCACGTCCA GCTGCGTGTG TCTCTGCCCC CTAGGGTCTC -4274
 GGGGTTTTTA TAGGCATAGG ACGGGGCGT GGTGGGCCAG GCGGCTCTTG GGAATGCAA CATTGGGTG -4204
 TGAAAGTAGG AGTGCTGTC CTCACCTAGG TCCACGGGCA CAGGCTGGG GATGGAGCCC CCGCCAGGGA -4134
 CCCGCCCTTC TCTGCCAGC ACTTTCCTG CTGCCACACA GAGTGGCAGT TTCCACAAGC -4064
 ACTAAGCATC CTCTTCCCA AAGACCCAGC ATTGACACC CTGGACATT GCCCACAGC CCTGGGAATT -3994

 c-Myc
 CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCG CTGTTTTATT TTAATAGCTA -3924
 CAAAGCAGGG AAATCCCTGC TAAATGTCC TTTAACAAAC TGTTAAACA AACGGGTCCA TCCGCACGGT -3854
 GGACAGTTCC TCACAGTGAA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG GAATTACGCT -3784
 GAGTCAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAAG AAAAGAATTC ACCCATGGC -3714
 AGGGGAGTGG TTAGGGGGGT TAAGGACCGT GGGGGCGGCA GCTGGGGGCT ACTGCACGCA CCTTTTACTA -3644
 AAGCCAGTTT CTTGGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG TGGGGATGGG -3574
 GGAACCCGGA GGCTGTGCCA TCTTTGCCAT GCCCGAGTGT CTTGGGCAGG ATAATGCTCT AGAGATGCCC -3504
 ACGTCTGAT TCCCCAAAC CTGTGGACAG AACC CGCCG GCGCCAGGGC CTTTGCAGGT GTGATCTCCG -3434
 TGAGGACCTT GAGTCTGGG ATCCTTCGGG ACTACCTGCA GGCCGAAAA GTAATCCAGG GGTCTGGGA -3364
 AGAGGCGGGC AGGAGGGTCA GAGGGGGGCA GCCTCAGGAG GATGGAGGCA GTCAGTCTGA GGCTGAAAAG -3294
 GGAGGGAGGG CCTCGAGCCC AGGCCTGCAA GCGCCTCCAG AAGCTGGAAA AAGCGGGGAA GGGACCCTCC -3224
 ACGGAGCCTG CAGCAGGAAG GCACGGCTGG CCCTTAGCCC ACCAGGGCCC ATCGTGGACC TCCGGCCTCC -3154
 GTGCCATAGG AGGGCACTCG CGTGCCTT CTAGCATGAA GTGTGTGGGG ATTTGCAGAA GCAACAGGAA -3084
 ACCCATGCAC TGTGAATCTA GGATTATTTT AAAACAAAGG TTTACAGAAA CATCCAAGGA CAGGGCTGAA -3014
 GTGCTCCCG GCAAGGGCAG GGCAGGCAG AGTGATTTTA TTAGCTATT TTATTTTATT TACTTACTTT -2944
 CTGAGACAGA GTTATGCTCT TGTGCCCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCAACCTC -2874
 CGTCTCCTGG GTTCAAGCAA TTCTCGTGCC TCAGCCTCCC AAGTAGCTGG GATTTTCAGGC GTGCACCACC -2804
 ACACCCGGCT AATTTTGTAT TTTTAGTAGA GATGGGCTTT CACCATGTTG GTCAAGCTGA TCTCAAATC -2734
 CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT -2664
 GGCCTATTTA ACCATTTTAA AACTTCCTG GGCTCAAGTC ACACCACTG GTAAGGAGTT CATGGAGTTT -2594
 AATTTCCCT TACTCAGGA GTTACCTCC TTTGATATT TCTGTAATTC TTCGTAGACT GGGGATACAC -2524
 CGTCTCTGA CATATTCACA GTTCTGTGA CCACCTGTTA TCCCATGGGA CCCACTGCAG GGGCAGCTGG -2454
 GAGGCTGCAG GCTTCAGGTC CCAGTGGGGT TGCCATCTGC CAGTAGAAAC CTGATGTAGA ATCAGGGCGC -2384
 AAGTGTGGAC ACTGCTCTGA ATCTCAATGT CTCAGTGTGT GCTGAAACAT GTAGAAATTA AAGTCCATCC -2314
 CTCCTACTCT ACTGGGATTG AGCCCTTCC CTATCCCCC CCAGGGGCG AGGAGTTCCT CTCCTCCTG -2244
 TGGAGGAAG AATGATACTT TGTATTTTT CACTGCTGGT ACTGAATCCA CTGTTTCATT TGTGTTTG -2174
 TTTGTTTTGT TTTGAGAGGC GGTTCCTC TTTGTGCTCA GGCTGGAGG AGTGCAATGG CGCATCTTG -2104
 GCTTACTGCA GCCTCTGCCT CCCAGGTTCA AGTGATTCTC CTGCTTCCG CTCCATTG GCTGGGATTA -2034
 CAGGACCCG CCACCATGCC CAGCTAATT TTTGTATTT TAGTAGAGAC GGGGGTGGG GGGGTTACC -1964

3

Fig.: 11



SEQUENZPROTOKOLL

<110> Bayer AG

5 <120> Regulatorische DNA-Sequenzen aus der 5i-Region vom Gen
der humanen katalytischen Telomerase-Untereinheit und
deren diagnostische und therapeutische Verwendung

10 <130> LeA32805-Ausland

<140>

<141>

15 <160> 20

<170> PatentIn Vers. 2.0

<210> 1

20 <211> 5126

<212> DNA

<213> Homo sapiens

<400> 1

25 gagctctgaa ccgtggaac gaacatgacc cttgcctgcc tgcttccctg ggtgggtcaa 60
gggtaatagaa gtgggtgtgca ggaaatggcc atgtaaatta cacgactctg ctgatgggga 120
ccgttccttc catcattatt catcttcacc cccaaggact gaatgattcc agcaacttct 180
tcgggtgtga caagccatga caaaactcag tacaacacacc actcttttac taggccacaca 240
gagcacgggc cacaccctg atatattaag agtccaggag agatgaggct gctttcagcc 300
accaggctgg ggtgacaaca gcggctgaac agtctgttcc tctagactag tagaccctgg 360
30 caggcactcc cccaaattct agggcctggg tgctgcttcc cgaggcgcc atctgccctg 420
gagactcagc ctgggggtgcc acactgaggc cagccctgtc tccacaccct ccgcctccag 480
gcctcagctt ctccagcagc ttccctaaacc ctgggtgggc cgtgttccag cgctactgtc 540
tcacctgtcc cactgtgtct tgtctcagcg acgtagctcg cacggttctt cctcacatgg 600
ggtgtctgtc tcttcccca acactcacat gcgttgaagg gaggagattc tgcgcctccc 660
35 agactggctc ctctgagcct gaacctggct cgtggccccc gatgcagggt cctggcgctc 720
ggctgcacgc tgacctccat ttccaggcgc tccccgtctc ctgtcatctg ccggggcctg 780
ccggtgtgtt ctctgtttc tgtgtcctt tccacgtcca gctgcgtgtg tctctgcccg 840
ctagggtctc ggggttttta taggcatagg acggggcgct ggtgggcccag ggcgctcttg 900
ggaaatgcaa catttggttg tgaaagttag agtgctgtc ctcacctagg tccacgggca 960
40 caggcctggg gatggagccc ccgccaggga cccgcccttc tctgcccagc actttcctgc 1020
ccccctccct ctggaacaca gagtggcagt ttccacaagc actaagcatc ctcttcccaa 1080
aagaccacgc attggcaccc ctggacattt gccccacagc cctgggaatt cacgtgacta 1140
cgacatcat gtacacactc ccgtccacga ccgacccccg ctgttttatt ttaatagta 1200
caaaagcagg aaatccctgc taaaatgccc tttaacaaac tgggttaaaca aacgggtcca 1260
45 tccgcacggt ggacagtcc tcacagtga gaggaacatg ccgtttataa agcctgcagg 1320
catctcaagg gaattacgct gagtcaaaac tgccacctcc atgggatacg tacgcaacat 1380
gctcaaaaag aaagaatttc accccatggc aggggagtggt ttaggggggt taaggacggt 1440
gggggaggca gctgggggct actgcacgca ctttttacta aagccagttt cctgggtctg 1500
atggtatttg ctctgttatg ggagactaac cataggggag tggggatggg ggaaccggga 1560
50 ggctgtgcca tctttgccat gcccgagtgt cctgggcagg ataagtctct agagatgcc 1620
acgtcctgat tccccaaac ctgtggacag aaccgcggcg gccccaggcg ctttgaggt 1680
gtgatctccg tgaggaccct gaggtctggg atccttcggg actacctgca ggccccgaaa 1740
gtaatccagg ggttctggga agaggcgggc agggagggtca gaggggggca gcctcaggac 1800
gatggaggca gtcagtctga ggtgaaaag ggaggagggt cctcgagccc aggcctgcaa 1860
55 gcgcctccag aagctggaaa aagcggggaa gggaccctcc acggagcctg cagcagggaag 1920
gcacggctgg cccttagccc accaggggcc atcgtggacc tccggcctcc gtgccatagg 1980
agggcactcg cgctgccctt ctagcatgaa gtgtgtgggg atttgagaa gcaacaggaa 2040
acctatgcac tgtgaactta ggattatttc aaaaacaaagg ttacagaaa catccaagga 2100
cagggtgaa gtgcctccgg gcaagggcag ggcaggcacg agtgatttta tttagctatt 2160
60 ttattttatt tacttacttt ctgagacaga gttatgctct tgttgcccag gctggagtgc 2220
agcggcatga tcttggtcga ctgcaacctc cgtctcctgg gttcaagcaa ttctcgtgcc 2280
tcagcctccc aagtagctgg gatttcaggc gtgcaccacc acaccggctt aattttgtat 2340
ttttagtaga gatgggcttt caccatgttg gtcaagctga tctcaaaatc ctgacctcag 2400
gtgatccgcc cacctcagcc tcccaaagtg ctgggattac aggcatagag cactgcacct 2460

5 ggccctattta accatttttaa aacttccctg ggetcaagtc acacccactg gtaaggagtt 2520
 catggagttc aatttcccct ttactcagga gttaccctcc tttgatattt tctgtaattc 2580
 ttcgtagact ggggatacac cgtctcttga catattcaca gtttctgtga ccacctgtta 2640
 tcccattgga cccactgcag gggcagctgg gaggctgcag gcttcaggtc ccagtggggt 2700
 tgccatctgc cagtagaaac ctgatgtaga atcaggggcg aagtgtggac actgtcctga 2760
 atctcaatgt ctacgtgtgt gctgaaacat gtagaaatta aagtcctacc ctccactctt 2820
 actgggattg agcccttccc ctatccccc ccaggggcag aggagttcct ctactcctg 2880
 tggaggaagg aatgatactt tgttattttt cactgctggt actgaatcca ctgtttcatt 2940
 10 tgttggtttg tttgttttgt tttgagaggc ggtttcactc ttgttgctca ggctggaggg 3000
 agtgcaatgg cgcgatcttg gcttactgca gcctctgcct cccaggttca agtgattctc 3060
 ctgcttccgc ctcccatttg gctgggatta caggcaccgc ccaccatgcc cagctaattt 3120
 tttgtatttt tagtagagac ggggttgagg ggggttcacc atgttggtcca ggctgggtctc 3180
 gaacttctgt cctcagatga tccacctgcc tctgcctcct aaagtgcctg gattacaggt 3240
 15 gtgagccacc atgcccagct cagaatttac tctgtttaga aacatctggg tctgaggtag 3300
 gaagctcacc cactcaagt gttgtggtgt ttttaagcaa tgatagaatt tttttattgt 3360
 tgttagaaca ctcttgatgt tttacactgt gatgactaag acatcatcag cttttcaaag 3420
 acacactaac tgcacccata atactggggg gtcttctggg tatcagcaat cttcattgaa 3480
 tgccgggagg cgtttcctcg ccatgcacat ggtgttaatt actccagcat aatcttctgc 3540
 20 ttcatttctt tctcttccc cttttaaaat tgtgttttct atgttggtct ctctgcagag 3600
 aaccagtgtg agctacaact taacttttgt tggaaacaaat tttccaaacc gcccttttgc 3660
 cctagtggca gagacaattc acaaacacag ccctttaaaa aggccttaggg atcactaagg 3720
 ggattttctag aagagcgacc tgtaatccta agtatttaca agacgaggct aacctccagc 3780
 gagcgtgaca gcccaggagg ggtgcgaggc ctgttcaaat gctagctcca taaataaagc 3840
 25 aatttctctc ggagtttctt gaaagtagga aaggttacat ttaaggttgc gtttggttagc 3900
 atttctagtg ttgcccagct cagctacagc atccctgcaa ggcctcggga gaccagaagc 3960
 tttctcgcct ccttagatcc aaacttgagc aaccggaggt ctggattcct gggaagtcct 4020
 cagctgtcct gcggttgtgc cggggcccca ggtctggagg ggaccagtgg ccgtgtggct 4080
 tctactgtcg ggctggaagt cgggcctcct agctctgcag tccgaggctt ggagccaggt 4140
 30 gcctggaccc cgaggctgcc ctccaccctg tgcgggcggg atgtgaccag atgttgccct 4200
 catctgccag acagagtgcc ggggcccagg gtcaaggccg ttgtggctgg tgtgaggcgc 4260
 ccggtgcgcg gccagcagga gcgcctggct ccatttccca ccctttctcg acgggaccgc 4320
 cccggtgggt gattaacaga tttggggtgg tttgctcatg gtggggaccc ctgcgcctc 4380
 gagaacctgc aaagagaaat gacgggcctg tgtcaaggag cccaagtcgc ggggaagtgt 4440
 35 tgcagggagg cactccggga ggtcccgcgt gcccgtccag ggagcaatgc gtcctcgggt 4500
 tcgtccccag ccgcgtctac gcgcctcctg cctccccttc acgtccggca ttcgtggtgc 4560
 ccggagcccg acgccccgcg tccggacctg gaggcagccc tgggtctccg gatcaggcca 4620
 gcggccaaag ggtgcgcgca cgcacctgtt cccagggcct ccacatcatg gccctccct 4680
 cgggttacc cagagcctag gccgattcga cctctctcct ctggggccct cgctggcgtc 4740
 40 cctgcacctt gggagcgcga gcgcgcgcg ggcggggaag cgcgggccag acccccgggt 4800
 ccgcccggag cagctgcgct gtcggggcca ggccgggctc ccagtggatt cgcgggcaca 4860
 gacgcccagg accgcgctcc ccacgtggcg gagggactgg ggacccgggc acccgtctcg 4920
 ccccttacc ttcagctcc gcctcctccg cgcggacccc gcccctctcc gaccctccc 4980
 ggggtccccg cccagcccc tccgggcccct cccagcccct ccccttctct tccgcccgc 5040
 45 cgccctctcc tcgcggcgcg agtttcaggc agcgtgcgt cctgctgcgc acgtgggaag 5100
 ccctggcccc ggccaccccc gcgatg 5126

<210> 2

<211> 4042

<212> DNA

50

<213> Homo sapiens

<400> 2

55 gtttcaggca gcgctgcgtc ctgctgcgca cgtgggaagc cctggccccg gccacccccg 60
 cgatgccgcg cgtccccgc tgcgagccg tgcgctccct gctgcgcagc cactaccgcg 120
 aggtgctgcc gctggccacg ttcgtgcggc gcctggggcc ccagggtgtg cggctggtgc 180
 agcgcgggga cccggcggtt ttccgcgcgc tgggtggcca gtgcctggtg tgcgtgccct 240
 gggacgcagc gccgcccccc gccgccccct ccttccgcca ggtgtcctgc ctgaaggagc 300
 tgggtggccc agtgctgcag aggtctgtcg agcgcggcgc gaagaacgtg ctggccttcg 360
 60 gcttcgcgct gctggacggg gcccgcgggg gccccccga ggcccttacc accagcgtgc 420
 gcagctacct gcccaacacg gtgaccgacg cactgcgggg gagcggggcg tgggggctgc 480
 tgctgcgcgg cgtgggagac gacgtgctgg ttcacctgct ggacgcgtgc gcgctctttg 540
 tgctgggtgg tccagctgc gcctaccagg tgtgcggggc gccgctgtac cagctcggcg 600
 ctgccactca ggcccggccc ccgccacacg ctagtggacc ccgaaggcgt ctgggatgcg 660
 aacgggcctg gaaccatagc gtcagggagg ccggggctcc cctgggcctg ccagccccgg 720
 65 gtgcgaggag gcgcgggggc agtgcacgcc gaagtctgcc gttgcccaag aggccaggc 780

3 / 18

5 gtggcgctgc ccctgagccg gagcggacgc ccgttgggca ggggtcctgg gccaccccg 840
 gcaggacgcg tggaccgagt gaccgtggtt tctgtgtggt gtcacctgcc agaccgcgg 900
 aagaagccac ctctttggag ggtgcgctct ctggcagcg ccactccac ccactccgtg 960
 gccgcccagca ccacgcgggc ccccatcca catcgcgcc accacgtccc tgggacacgc 1020
 cttgtcccc ggtgtacgcc gagaccaagc acttctcta ctctcaggc gacaaggagc 1080
 agctgcggcc ctcttctcta ctacgtctc tgaggcccag cctgactggc gctcggaggc 1140
 tcgtggagac catctttctg ggttccaggc cctggatgcc agggactccc cgcagggttg 1200
 cccgcctgcc ccagcgctac tggcaaatgc ggccctgtt tctggagctg cttgggaacc 1260
 acgcgcagtg cccctacggg gtgtctctca agacgcactg cccgtgcca gctgcggtca 1320
 10 cccagcagc cggtgtctgt gcccgggaga agccccagg ctctgtggcg gccccgagg 1380
 aggaggacac agacccccgt cgctgtgctc agctgctccg ccagcacagc agccccctgg 1440
 aggtgtacgg ctctgtcgcg gcctgcctgc gccggtggt gccccaggc ctctggggct 1500
 ccaggcaca cgaacgcgc ttctcaggc acaccaagaa gttcatctc ctggggaagc 1560
 atgccaaagt ctcgctcgag gagctgacgt ggaagatgag cgtgcgggac tgcgcttg 1620
 15 tgcgcaggag cccagggtt ggtgtgttc cggccgcaga gcacctctg cgtgaggaga 1680
 tcttgcccaa gttctcgac tggctgatga gtgtgtacgt cgtcgagctg ctgaggtctt 1740
 tcttttatgt caggagacc acgtttcaaa agaaccaggct cttttctac cgggaagagt 1800
 tctggagcaa gttgcaagc attggaatca gacagcactt gaagagggtg cagctgcggg 1860
 agctgtcggg agcagaggtc aggcagcatc gggaagccag gccccctg ctgacgtcca 1920
 20 gactccgctt catcccaag cctgacgggc tgcggccgat tgtgaacatg gactacgtcg 1980
 tgggagccag aacgttccgc agagaaaaga gggccgagcg tctcacctcg aggggtgaag 2040
 cactgttcag cgtgtcaac tacgagcggg cgcgcgccc cggtctctg ggcgctctg 2100
 tgctgggctt ggacgatata cacaggcctt ggcgcacctt cgtgtgctg gtgccccg 2160
 25 aggaccgccc gcctgagctg tactttgtca aggtggatgt gacgggcgag tacgacacca 2220
 tccccagga caggctcagc gaggtcatcg ccagcatcat caaacccag aacacgtact 2280
 gcgtgcgtcg gtatgccgtg gtccagaagg ccgcccattg gcacgtccgc aaggccttca 2340
 agagccacgt ctctacctg acagacctcc agccgtacat gcgacagttc gtggctcacc 2400
 tgcaggagac cagcccgtg agggatgccg tcgtcatcga gcagagctcc tccctgaatg 2460
 30 agccagcagc tggcctcttc gacgtcttcc tacgttctat gtgccaccac gccgtgcgca 2520
 tgggggcaa gtcctacgtc cagtgccagg ggtcccgca gggctccatc ctctccacgc 2580
 tgctctgcag cctgtgtac ggcgacatgg agaacaagct gtttcggggg attcggcggg 2640
 acgggctgct cctgcgtttg gtggatgatt tcttgttggg gacacctcac ctacccacg 2700
 cgaaaacctt cctcaggacc ctggtccgag gtgtccctga gtatggctgc gtggtgaact 2760
 35 tgcggaagac agtgggtaac ttccctgtag aagacgaggc cctgggtggc acggcttttg 2820
 ttcagatgcc ggcccacggc ctattccctt ggtgcggcct gctgtggat acccgagcc 2880
 tgaaggtgca gagcgactac tccagctatg cccggacctc catcagagcc agtctacct 2940
 tcaaccgagg cttcaaggct gggaggaaca tgcgtcgcaa actctttggg gtcttgccgc 3000
 tgaagtgtca cagcctgtt ctggatttgc aggtgaacag cctccagacg gtgtgcacca 3060
 40 acatctacaa gatectctg ctgcaggcgt acaggtttca cgcattgtgt ctgcagctcc 3120
 catttcatca gcaagtttgg aagaacccca catttttctt gcgcgtcatc tctgacacgg 3180
 cctccctctg ctactccatc ctgaaagcca agaaccgagg gatgtcgtg ggggccaagg 3240
 tgcgcccggg cctctgtccc tccgaggcgg tgcagtggct gtgccaccaa gcattctctg 3300
 gcaagctgac tcgacaccgt gtcacctcct tgccactcct ggggtcactc aggacagccc 3360
 45 agacgcagct gagtcggaag ctcccgggga cgacgtgac tgccctggag gccgcagcca 3420
 acccggcact gccctcagac ttcaagacca tctggactg atggccacc gccacagcc 3480
 aggcgagag cagacaccag cagccctgtc acgcccggct ctacgtccca gggaggagg 3540
 ggcgcccac acccaggccc gcaccgctgg gactctgagg cctgagttag tgtttggcg 3600
 aggcctgcat gtccggtga aggtgagtg tccggtgag gcctgagcga gtgtccagcc 3660
 50 aagggtgtag tgtccagcac acctgcgtc ttacttccc cacaggctg cgctcggctc 3720
 caccacagg ccagcttttc ctaccaggga gcccggttc cactccccac ataggaatag 3780
 tccatcccca gatcgccat tgttaccctc tcgcccgtgc ctcttttgc ttccaccccc 3840
 accatccagg tggagaccct gagaaggacc ctgggagctc tgggaatttg gactgaccaa 3900
 aggtgtgccc tgtacacagg caggagacct gcacctggat gggggctcct gtgggtcaaa 3960
 55 ttgggggggag gtgctgtggg agtaaaatac tgaatatatg agtttttcag ttttgaaaaa 4020
 aaaaaaaaaa aaaaaaaaaa aa 4042

<210> 3

<211> 11276

<212> DNA

<213> Homo sapiens

<400> 3

acttgagccc aagagttcaa ggctacgggt agccatgatt gcaacaccac acgccagcct 60
 tgggtgacaga atgagaccct gtctcaaaaa aaaaaaaaaa aattgaaata atataaagca 120
 65 tcttctcttg ccacagtggg acaaaaccag aaatcaacaa caaggagaa tttgaaaact 180

| | | | | | | | |
|----|-------------|-------------|-------------|------------|------------|-------------|------|
| | atacaaacac | atgaaaaatta | aacaatatat | ttctgaatga | ccagtggatc | aatgaagaaa | 240 |
| | ttaaaaagga | aattgaaaaa | tttattttaag | caaatgataa | cggaaacata | acctctcaaa | 300 |
| | acccacggta | tacagcaaaa | gcagtgtctaa | gaaggaagtt | tatagctata | agcagctaca | 360 |
| 5 | tcaaaaaagt | agaaaaagcca | ggcgacgtgg | ctcatgcctg | taatcccagc | actttgggag | 420 |
| | gccaaggcgg | gcagatcgcc | tgaggtcagg | agttcggagc | cagcctgacc | aacacagaga | 480 |
| | aaccttgtcg | ctactaaaaa | tacaaaatta | gctgggcatg | gtggcacatg | cctgtaatcc | 540 |
| | cagctactcg | ggaggctgag | gcaggataac | cgcttgaacc | caggagggtg | aggttgcggt | 600 |
| | gagccgggat | tgcgccattg | gactccagcc | tgggtaacaa | gagtgaacc | ctgtctcaag | 660 |
| 10 | aaaaaaaaa | aagttagaaa | acttaaaaat | acaaccta | gatgcacctt | aaagaactag | 720 |
| | aaaagcaaga | gcaaaactaa | cctaaaattg | gtaaaagaaa | agaaataata | aagatcagag | 780 |
| | cagaaataaa | tgaactgaa | agataacaat | acaaaagatc | aacaaaatta | aaagtgtggt | 840 |
| | ttttgaaaag | ataaacaaaa | ttgacaaacc | tttgcccaga | ctaagaaaaa | aggaaaagaa | 900 |
| | acctaataaa | ataaagtcag | agatgaaaaa | agagacatta | caactgatac | cacagaaatt | 960 |
| 15 | caaaggatca | ctagaggcta | ctatgagcaa | ctgtacacta | ataaattgaa | aaacctagaa | 1020 |
| | aaaatagata | aattcctaga | tgcatacaac | ctaccaagat | tgaacctatg | agaaatccaa | 1080 |
| | agcccaaca | gaccaataac | aataatggga | ttaaagccat | aataaaaagt | ctcctagcaa | 1140 |
| | agagaagccc | aggacccaat | ggcttccttg | ctggatttta | ccaatcattt | aaagaagaa | 1200 |
| | gaattccaat | cctactcaaa | ctattctgaa | aaatagagga | aagaataact | ccaaactcat | 1260 |
| 20 | tctacatggc | cagtattacc | ctgattccaa | aaccagacaa | aaacacatca | aaaacaacaa | 1320 |
| | aacaaaaaaa | cagaaagaaa | gaaaactaca | ggccaatatc | cctgatgaat | actgatacaa | 1380 |
| | aaatcctcaa | caaaacacta | gcaaaccaaa | ttaaacacaa | ccttcgaaag | atcattcatt | 1440 |
| | gtgatcaagt | gggatttatt | ccagggatgg | aaggatggtt | caacatatgc | aaatcaatca | 1500 |
| | atgtgatata | tcacccaac | aaaatgaagt | acaaaaacta | tatgattatt | tcactttatg | 1560 |
| 25 | cagaaaaagc | atgtgataaa | attctgcacc | cttcagtata | aaaaacccca | aaaaaacagg | 1620 |
| | tatacaagaa | acatacaggc | caggcacagt | ggctcacacc | tgcgatccca | gcactctggg | 1680 |
| | aggccaaggt | gggatgattg | cttgggcccc | ggagtgtgag | actagcctgg | gcaacaaaa | 1740 |
| | gagacctggt | ctacaaaaaa | cttttttaaa | aaattagcca | ggcatgatgg | catatgcctg | 1800 |
| | tagtcccagc | tagtctggag | gctgaggtgg | gagaatcact | taagcctagg | aggtcagagg | 1860 |
| 30 | tgcagtggc | catgaacatg | tcactgtact | ccagcctaga | caacagaaca | agaccccat | 1920 |
| | gaataagaag | aaggagaagg | agaaggga | agggaggag | aaggaggag | gaggagaagg | 1980 |
| | aggaggtgga | ggagaagtgg | aaggggaa | ggaaggga | gagggaagaa | aagaaacata | 2040 |
| | tttcaacata | ataaaagccc | tatatgacag | accgaggtag | tattatgagg | aaaaactgaa | 2100 |
| | agcctttcct | ctaagatctg | gaaaatgaca | agggccact | ttcaccactg | tgattcaaca | 2160 |
| 35 | tagtactaga | agtcctagct | agagcaatca | gataagagaa | agaaataaaa | ggcatccaaa | 2220 |
| | ctggaaagga | agaagtcaaa | ttatcctgtt | tgcagatgat | atgatcttat | atctggaaaa | 2280 |
| | gacttaagac | accactaaaa | aactattaga | gctgaaattt | ggtacagcag | gatacaaaa | 2340 |
| | caatgtacaa | aaatcagtag | tatttctata | ttccaacagc | aaacaatctg | aaaaagaac | 2400 |
| 40 | caaaaaagca | gctacaaata | aaataaaca | gctaggaatt | aaccaagaa | gtgaaagatc | 2460 |
| | tctacaatga | aaactataaa | atgttgataa | aagaaattga | agagggcaca | aaaaaagaaa | 2520 |
| | agatattcca | tgttcataga | ttggaagaat | aaatactgtt | aaaatgtcca | tactacccaa | 2580 |
| | agcaattttac | aaattcaatg | caatccctat | taaaatacta | atgacgttct | tcacagaaat | 2640 |
| | agaagaaaca | atttctaagt | ttgtacagaa | ccacaaaaga | cccagaatag | ccaaagctat | 2700 |
| | cctgaccaa | aagaacaaaa | ctggaagcat | caattacct | gacttcaaat | tatactacaa | 2760 |
| 45 | agctatagta | acccaaacta | catggtactg | gcataaaaa | agatgagaca | tggaccagag | 2820 |
| | gaacagaata | gagaatccag | aaacaaatcc | atgcatctac | agtgaactca | tttttgacaa | 2880 |
| | aggtgccaa | aacatacttt | ggggaaaaaga | taatctcttc | aataaatggt | gctggaggaa | 2940 |
| | ctggatatcc | atatgcaaaa | taacaatact | agaactctgt | ctctcaccat | atacaaaagc | 3000 |
| | aaatcaaaat | ggatgaaagg | cttaaatcta | aaacctcaaa | ctttgcaact | actaaaagaa | 3060 |
| 50 | aacaccggag | aaactctcca | ggacattgga | gtgggcaaa | acttcttgag | taattccctg | 3120 |
| | caggcacag | caaccaaaagc | aaaaacagac | aaatgggatc | atatcaagtt | aaaaagcttc | 3180 |
| | tgcccagcaa | aggaaacaat | caacaaagag | aagagacaac | ccacagaatg | ggagaatata | 3240 |
| | tttgcaaaact | attcatctaa | caaggaatta | ataaccagta | tatataagga | gctcaaaacta | 3300 |
| | ctctataaga | aaaacaccta | ataagctgat | tttcaaaaat | aagcaaaaga | tctgggtaga | 3360 |
| 55 | cattttctcaa | aataagtcac | acaaatggca | aacaggcatc | tgaaaatgtg | ctcaacacca | 3420 |
| | ctgatcatca | gagaaatgca | aatcaaaact | actatgagag | atcatctcat | cccagttaaa | 3480 |
| | atggctttta | ttcaaaagac | aggcaataac | aaatgccagt | gaggatgtgg | ataaaaggaa | 3540 |
| | acccttggac | actgttgggtg | ggaatggaaa | ttgctaccac | tatggagaac | agtttgaaag | 3600 |
| | ttcttcaaaa | aactaaaaat | aaagctacca | tacagcaatc | ccattgctag | gtatatctc | 3660 |
| 60 | caaaaaaggg | aatcagtgtg | tcacaaagct | atctccactc | ccacatttac | tgcagcactg | 3720 |
| | ttcatagcag | ccaaggtttg | gaagcaacct | cagtgtccat | caacagacga | atggaaaaag | 3780 |
| | aaaatgtggt | gcacatacac | aatggagtac | tacgcagcca | taaaaaagaa | tgagatcctg | 3840 |
| | tcagttgcaa | cagcatgggg | ggcactggtc | agtatgttaa | gtgaaataag | ccaggcacag | 3900 |
| | aaagacaaac | ttttcatggt | ctcccttact | tgtgggagca | aaaattaaaa | caattgacac | 3960 |
| 65 | agaaatagag | gagaatgggtg | gttctagagg | gggtggggac | agggtgacta | gagtcaacaa | 4020 |
| | taattttattg | tatgttttaa | aataactaaa | agagtataat | tgggttgttt | gtaacacaaa | 4080 |

5
10
15
20
25
30
35
40
45
50
55
60
65

gaaaggataa atgcttgaag gtgacagata cccatttac cctgatgtga ttattacaca 4140
ttgtatgcct gtatcaaaat atctcatgta tgctatagat ataaacccta ctatattaaa 4200
aattaaaatt ttaattggcca ggcacggtgg ctcatgtccg taatcccagc accttggggag 4260
gccgaggcgg gtggatcacc tgaggtcagg agtttgaac cagtctggcc accatgatga 4320
aaccctgtct ctactaaaga tacaaaaatt agccaggcgt ggtggcacat acctgtagtc 4380
ccaactactc aggaggctga gacaggagaa ttgcttgaac ctgggaggcg gaggttgcag 4440
tgagccgaga tcatgccact gcactgcagc ctgggtgaca gagcaagact ccatctcaaa 4500
acaaaaacaa aaaaaagaag attaaaattg taatttttat gtaccgtata aatatatact 4560
ctactatatt agaagttaaa aattaaaaca attataaaag gtaattaacc acttaattcta 4620
aaataagaac aatgtatgtg gggtttctag cttctgaaga agtaaaagt atggccacga 4680
tggcagaaat gtgaggaggg aacagtggaa gttactgttg ttagacgctc atactctctc 4740
taagtgtact aattttaacc aaagacaggc tgggagaagt taaagaggca ttctataagc 4800
cctaaaacaa ctgctaataa tgggtgaaagg taatctctat taattaccaa taattacaga 4860
tatctctaaa atcgagctgc agaattggca cgtctgatca caccgtcctc tcattcacgg 4920
tgcttttttt cttgtgtgct tggagatttt cgattgtgtg ttcgtgtttg gttaaactta 4980
atctgtatga atctgaaac gaaaaatggt ggtgatttcc tccagaagaa ttagagtacc 5040
tggcaggaag cagggtggctc tgtggacctg agccacttca atcttcaagg gtctctggcc 5100
aagacccagg tgcaaggcag aggcctgatg acccgaggac aggaaaagctc ggatgggaaag 5160
gggcatgag aagcctgcct cgttgggtgag cagcgcatga agtgccctta ttacgcttt 5220
gcaaagattg ctctggatac catctggaaa aggcggccag cggaatgca aggagtcaga 5280
agcctcctgc tcaaaaccag gccagcagct atggcgccca cccggcgctg tgccagaggg 5340
agaggagtca aggcacctcg aagtatggct taaatcttt tttcacctga agcagtgacc 5400
aaggtgtatt ctgagggaag cttgagttag gtgccttctt taaaacagaa agtcatggaa 5460
gcacccctct caagggaana ccagacgccc gctctgcggt catctacctc ttctctctc 5520
ccctctcttg ccctcgcggt ttctgatcgg gacagagtga ccccggtgga gcttctccga 5580
gcccggtgctg aggacctctc tgcaaaagggc tccacagacc cccgcctgga agagaggagt 5640
ctgagcctgg cttaataaca aactgggatg tggctggggg cggacagcga cggcgggatt 5700
caaagactta attccatgag taaattcaac ctttccacat ccgaatggat ttggatttta 5760
tcttaatttt ttcttaaaat tcatcaaaata acattcagga ctgcagaaat ccaaggcgt 5820
aaaacaggaa ctgagctatg ttgccaagg tccaaggact taataacat gttcagaggg 5880
atcttctgcc ctaagtactt ttatttggt ttcataaggt ggcttagggg gcaagggaag 5940
gtacacgagg agaggcctgg ggcgcagggc tatgagcacg gcagggccac cgggggagaga 6000
gtccccggcc tgggaggctg acagcaggac cactgaccgt cctccctggg agctgccaca 6060
ttgggcaacg cgaaggcggc cacgtgctg gtgactcagg accccatacc ggcttctctg 6120
gcccaccac actaaccag gaagtacagg agctctgaac ccgtggaac gaacatgacc 6180
cttgccctgcc tgcttccctg ggtgggtcaa gggtaatgaa gtgggtgtga ggaatggcc 6240
atgtaatta cagactctg ctgatgggga ccgttctctc catcattatt catcttcac 6300
cccaaggact gaatgattcc agcaactctc tcgggtgtga caagccatga caaaactcag 6360
tacaacacc actcttttac taggcccaca gagcacggsc cacaccctg atatattaag 6420
agtccaggag agatgaggct gctttcagcc accaggctgg ggtgacaaca gcggctgaac 6480
agtctgttcc tctagactag tagaccctgg caggcactcc cccagattct agggcctggt 6540
tgctgcttcc cgaggcgcc atctgccctg gagactcagc ctgggtgacc aactgaggc 6600
cagccctgtc tccacacctc ccgctccag gcctcagct ctccagcagc ttcttaaac 6660
ctgggtgggc cgtgttccag cgctactgtc tcacctgtcc cactgtgtct tgtctcagc 6720
acgtagctcg caggttctct cctcacatct ggtgtctgtc tccctcccca acactcacat 6780
gcgttgaagg gaggagattc tgcgcctccc agactggctc ctctgagcct gaacctggct 6840
cgtggccccc gatgcagggt cctggcgctc ggtgcacgc tgacctccat ttccaggcgc 6900
tcccgtctc ctgtcatctg ccggggcctg ccgggtgtgt cttctgttcc tgtgtcctt 6960
tccacgtcca gctgcgtgtg tctctgcccc ctagggtctc ggggttttta taggcatagg 7020
acgggggctg ggtgggcccag ggcgctcttg ggaaatgcaa catttgggtg tgaagtagg 7080
agtgcctgtc ctacactagg tccacgggca caggcctggg gatggagccc ccgcccaggga 7140
ccgcctctc tctgcccagc actttcctgc cccctccct ctggaacaca gaggggcagt 7200
ttccacaagc actaagcatc ctcttcccaa aagaccagc attggcacc ctggacattt 7260
gccccacagc cctgggaatt cacgtgacta cgcacatcat gtacacactc ccgtccacga 7320
ccgacccccg ctgttttatt ttaatagcta caaagcagg aaatccctgc taaaatgtcc 7380
tttaacaac tggttaaaac aacgggtcca tccgcacggg ggacagtcc tcacagtga 7440
gaggaacatg ccgtttataa agcctgcagg catctcaagg gaattacgt gagtcaaac 7500
tgccacctcc atgggatacg tacgcaacat gctcaaaaag aaagaatttc acccatggc 7560
aggggagtg ttaggggggt taaggacggt gggggcgga gctgggggt actgcacgca 7620
ccttttacta aagccagttt cctggtctg atggtattg ctgagttatg ggagactaac 7680
cataggggag tggggatggg ggaacccgga ggctgtgcca tctttgccat gcccgagtgt 7740
cctgggcagg ataagtctct agagatgccc acgtcctgat tccccaaac ctgtggacag 7800
aaccgcgccg gcccagggc ctttgagggt gtgatctccg tgaggacctt gaggtctggg 7860
atccttcggg actacctgca ggccgaaaa gtaatccagg ggttctggga agaggcgggc 7920
aggagggtca gaggggggca gcctcaggac gatggaggca gtcagctctga ggctgaaaag 7980

6 / 18

5 ggagggaggg cctcgagccc aggcctgcaa ggcctccag aagctggaaa aagcggggaa 8040
 gggaccctcc acggagcctg cagcaggaag gcacggctgg cccttagccc accagggccc 8100
 atcgtggacc tccggcctcc gtgcatagg agggcactcg cgctgccctt ctagcatgaa 8160
 gtgtgtgggg atttgcagaa gcaacaggaa acccatgcac tgtgaatcta ggattatttc 8220
 aaacaaaagg ttacagaaa catccaagga cagggtgaa gtgcctccgg gcaagggcag 8280
 ggcaggcacg agtgatttta tttagctatt ttattttatt tacttacttt ctgagacaga 8340
 gttatgtctt tgtgcccag gctggagtgc agcggcatga tcttggtca ctgcaacctc 8400
 cgtctcctgg gttcaagcaa ttctcgtgcc tcagcctccc aagtagctgg gatttcaggc 8460
 10 gtgcaccacc acaccggct aattttgtat ttttagtaga gatgggcttt caccatgttg 8520
 gtcaagctga tctcaaatc ctgacctcag gtgatccgcc cactcagcc tcccaaagt 8580
 ctgggattac aggcagagc cactgcacct ggctatttta accattttaa aacttccctg 8640
 ggctcaagtc acaccactg gtaaggagtt catggagttc aatttccctt ttactcagga 8700
 gttaccctcc tttgatattt tctgtaattc ttcgtagact ggggatacac cgtctcttga 8760
 catattcaca gtttctgtga ccacctgtta tcccatggga ccactgcag gggcagctgg 8820
 15 gaggctgcag gcttcaggtc ccagtggggt tgccatctgc cagtagaaac ctgatgtaga 8880
 atcagggcgc aagtgtggac actgtcctga atctcaatgt ctcagtgtgt gctgaaacat 8940
 gtagaaatta aagtgcctcc ctctactct actgggattg agccccctcc ctatcccccc 9000
 ccaggggcag aggagttcct ctccactctg tggagggaagg aatgactctt tgttattttt 9060
 cactgctggt actgaatcca ctgtttcatt tgttggtttg tttgttttgt tttgagaggc 9120
 20 ggtttcactc ttgttgctca ggctggaggg agtgcaatgg cgcgatcttg gcttactgca 9180
 gcctctgcct cccaggttca agtgattctc ctgcttccgc ctcccatttg gctgggatta 9240
 caggcacccg ccaccatgcc cagctaattt tttgtatttt tagtagagac ggggggtgggt 9300
 ggggttcacc atgttggcca ggctggctct gaacttctga cctcagatga tccacctgcc 9360
 tctgcctcct aaagtgtcgg gattacaggt gtgagccacc atgccagct cagaatttac 9420
 25 tctgtttaga aacatctggg tctgaggtag gaagctcacc ccactcaagt gttgtgggtg 9480
 ttttaagcaa tgatagaatt ttttatttgt tgttagaaca ctcttgatgt ttactactgt 9540
 gatgactaag acatcatcag cttttcaaag acacactaac tgcaccata atactggggt 9600
 gtcttctggg tatcagcaat ctctattgaa tgccgggagg cgtttcctcg ccactgcacat 9660
 ggtgttaatt actccagcat aatcttctgc ttccatttct tctcttccct cttttaaaat 9720
 30 tgtgttttct atgttggctt ctctgcagag aaccagtgt agctacaact taacttttgt 9780
 tggaaacaaat tttccaaacc gcccttttgc cctagtggca gagacaattc acaaacacag 9840
 ccctttaaaa aggccttaggg atcactaagg ggatttctag aagagcgacc tghtaatcta 9900
 agtatattaca agacgaggct aacctccagc gagcgtgaca gccaggagg ggtgcgaggc 9960
 35 ctgttcaaat gctagctcca taaataaagc aatttcctcc ggcagtttct gaaagtagga 10020
 aaggttacat ttaagggtgc gtttgttagc atttcagtgt ttgccgacct cagctacagc 10080
 atccctgcaa ggcctcgga gaccagaag tttctcgccc ccttagatcc aaacttgagc 10140
 aaccggaggt ctggattcct gggaaatcct cagctgtcct gcggttgtgc cggggcccca 10200
 ggtctggagg ggaccagtgg ccgtgtggct tctactgctg ggctggaagt cgggcctcct 10260
 40 agctctgcag tccgaggctt ggagccaggt gcctggacc cgaggctgcc ctccacctg 10320
 tgcgggcggg atgtgaccag atgtggcct catctgccag acagagtgcc ggggccagg 10380
 gtcaaggccg ttgtggctgg tgtgaggcgc ccggtgcgcg gccagcagga gcgcctggct 10440
 ccatttccca ccctttctcg acgggaccgc ccggtgggt gattaacaga tttgggggtg 10500
 tttgtctatg gtggggaccc ctgcgcgctc gagaacctgc aaagagaaat gacggcgctg 10560
 45 tgtcaaggag ccaagtgcg ggggaagtgt tgcaggga cactccggga ggtcccgcgt 10620
 gccggtccag ggagcaatgc gtctcgggt tgcgtcccag ccgctctac gcgcctcgt 10680
 cctccccttc acgtccggca ttcgtggtgc ccggagcccg acgccccgcg tccggacctg 10740
 gaggcagccc tgggtctccg gatcaggcca gcggccaaag ggtcgccgca cgcacctgtt 10800
 cccagggcct ccacatcatg gccctccct cgggttacc caccagctag gccgattcga 10860
 50 cctctctccg ctggggccct cgctggcgte cctgcacct gggagcgcg ggcggcgcg 10920
 ggcggggaag cgcggccag accccgggt ccgcccggag cagctgcgct gtcggggcca 10980
 ggccgggctc ccagtggatt cgcgggcaca gacgccagg accgcgctcc ccacgtggcg 11040
 gagggactgg ggaccgggc acccgctctg ccccttcacc ttccagctcc gcctcctccg 11100
 cgcggacccc gcccgctccc gaccctccc ggggtcccg cccagcccc tccgggcct 11160
 55 cccagccct ccccttccct tcccgggccc cgcctctcc tcgcggcgc agtttcaggc 11220
 agcgtgcgt cctgctgcgc acgtgggaag ccctggcccc gcccaccccc gcgatg 11276

 <210> 4
 <211> 104
 <212> DNA
 60 <213> Homo sapiens

 <400> 4
 gtgggctcc ccggggtcgg cgtccggctg ggggtgaggg cggccggggg gaaccagcga 60
 catgcggaga gcagcgagg cgactcaggg cgcttcccc gcag 104
 65

<210> 5
<211> 8616
<212> DNA
<213> Homo sapiens

5

<400> 5

10

15

20

25

30

35

40

45

50

55

60

65

gtgaggaggt ggtggccgtc gaggggccag gccccagagc tgaatgcagt agggggtcag 60
aaaagggggc aggcagagcc ctggtcctcc tgtctccatc gtcacgtggg cacacgtggc 120
ttttcgctca ggacgtcgag tggacacggt gatctctgcc tctgctctcc ctctgtcca 180
gtttgcataa acttacgagg ttcaccttca cgttttgatg gacacgcggt ttccaggcgc 240
cgaggccaga gcagtgaaca gaggaggctg ggcgcggcag tggagccggg ttgccggcaa 300
tgggaggaag tgtctggaag cacagacgct ctggcgaggg tgcctgcagg ttacctataa 360
tcctcttcgc aatttcaagg gtgggaatga gagggtggga cgagaacccc ctcttctctg 420
gggtgggagg taagggtttt gcaggtgcac gtggtcagcc aatatgcagg tttgtgttta 480
agatttaatt gtgtgttgac ggcaggtgac ggtggctcac gccggtaatc ccagcacttt 540
gggaagctga ggcaggtgga tcacctgagg tcaggagttt gagaccagcc tgaccaacat 600
ggtgaaaccc tatctgtact aaaaatacaa aaattagctg ggcatggtag tgtgtgctg 660
taatcccagc tacttgggag gctgaggcag gagaatcact tgaaccagg aggcggaggc 720
tgagtgagc tgagattgtg ccattgtact ccagcctggg cgacaagagt gaaactctgt 780
ctttaaaaaa aaaaagtgtt cgttgattgt gccaggacag ggtagaggga gggagataag 840
actgttctcc agcacagatc ctggtcccat ctttaggtat gaagagggcc acatgggagc 900
agaggacagc agatggctcc acctgctgag gaaggacag tgtttgtggg tgttcagggg 960
atggtgctgc tgggcccctg cgtgtcccca ccctgtttt ctggatttga tgttaggaa 1020
cctccgctcc agccccctt tggctccagc tgcctccagg ccctaccgtg gcagctagaa 1080
gaagtcgga tttcaccccc tccccacaaa ctcccaagac atgtaagact tccggccatg 1140
cagacaagga gggtagacct cttggggctc tttttttct ttttttctt ttatgggtgc 1200
aaaagtcata taacatgaga ttggcactcc taacaccgtt ttctgtgtac agtcagaa 1260
tgctaactcg gcggtgttta cagcaggttg cttgaaatgc tgcgtcttgc gtgactggaa 1320
gtccctaccc atcgaacgyc agctgcctca cacctgctgc ggctcagggt gaccacgccg 1380
agtcagataa gcgtcatgca acccagtttt gctttttgtg ctccagcttc ctctgttag 1440
gagagtttga gttctctgat caggactctg cctgtcattg ctgttctctg acttcagatg 1500
aggtcacaat ctgcccctgg cttatgcagg gaggtaggag tgggtcccgg gtgtccctgt 1560
cacgtgcagg gtgagtgagg cgttgcctcc aggtgtccct gtcacgtgta gggtagtgta 1620
ggcgcgggcc ccgggtgtcc ctgtcccgtg cagcgtgatt gagggtgtgg ccccggtgtg 1680
ccctgtcacg ttaggggtga gtgaggcgcc atcccgggt gtccctgtca cgtgtagggt 1740
gagtgaggcg tgggtcccgg gtgtccctgt cccgtgcagg gtgagtgagg cactgtcccc 1800
gggtgtccct gtcacgtgca gggtagtgta ggcgcggctc ccgggtgtcc ctctcagggt 1860
taggggtgag gaggcgcgcc cccagggtgt cctgtcacg ttaggggtga gtgaggcacc 1920
gtccctgggt gtccctccca ggtatagggt gaggtaggca ctgtcccgg gtgtccctgt 1980
cacgtgcagg gtgagtgagg cgcggccccc ggggtgtccct ctccaggtgca gggtagtgta 2040
ggcgctgtcc ctgggtgtcc ctgtctctgt tagggtagt gaggctctgt ccccgaggtg 2100
ccttggtggt tgcctcactg agcttgcctc tgaatgtttg ctctttctat agccacagct 2160
gcgcgggttg cccattgcct gggtagatgg tgcaggcgca gtgctggtcc ccaagcctat 2220
cttttctgat gctcggctct tcttgggtcac ctctccgttc cattttgcta cggggacagc 2280
ggactgcagg ctctgcctc cgcggtgca ggcactgcag ccacagcttc aggtccgctt 2340
gcctctgttg ggcctggctt gctcaccacg tgcccggcac atgcatgctg ccaatactcc 2400
tctcccagct tgtctcatgc cgaggctgga ctctgggctg cctgtgtctg ctgccacgtg 2460
ttgctggaga catcccagaa aggtttctct gtgccctgaa ggaaagcaag tcacccagc 2520
cccctcactt gtccctgttt ctcccaagct gccctctgct tgggccccct tgggtgggtg 2580
gcaacgcttg tcaccttatt ctgggcacct gccgtcatt gcttaggctg ggctctgcct 2640
ccagtcgccc cctcacatgg attgacgtcc agccacagg tggagtgct ctgtctgtct 2700
cctgctctga gacccacgtg gagggccggt gtctccgcca gcctctgca gacttccctc 2760
ttgggtctta gttttgaatt tcaactgatt acctctgacg tttctatctc tccattgtat 2820
gctttttctt ggtttattct ttcattcctt ttctagcttc ttagtttagt catgccttct 2880
cctctaagtg ctgccttacc tgcacctgt gttttgatgt gaagtaatct caacatcagc 2940
cactttcaag tgttctttaa atacttcaaa gtgttaatac ttcttttaag tattcttatt 3000
ctgtgatttt tttctttgtg cacgtgtgt tttgacgtga aatcattttg atatcagta 3060
cttttaagta ttctttagct tattctgtga tttctttgag cagttagtta ttgaacatg 3120
gttttatgtc aagatatgta gagtatacag atacgtagag tatttttaag tatcatttta 3180
ttattgattt ctaactcagt tgtgtagtgg tctgtataat accaattatt tgaagtttgc 3240
ggagccttgc tttgtgatct agtgtgtgca tgggttccag aactgtccat tgtaaatttg 3300
acatccgtc aatagtgggc atgcatgttc actatatcca gcttattaag gtccagtgca 3360
aagcttctgt ctccctctag atgcatgaaa ttccaagaag gaggccatag tccctcacct 3420
gggggatggg tctgttcatt tcttctcgtt tggtagcatt tatgtgaggc attgttaggt 3480
gcatgcacgt ggtagaattt ttatcttctc gatgagtga tcttttggag acttctatgt 3540

| | | | | | | | |
|----|-------------|-------------|-------------|-------------|-------------|-------------|------|
| | ctctagtaat | ctagtaattc | tttttttaaa | ttgtctcttag | tactgccaca | ctgggcttct | 3600 |
| | tttgattagt | attttctctg | tgtgtctgtt | ttctgccttt | aatttatata | tatatatata | 3660 |
| | tttttttttt | ttttgagaca | gagtccttgg | ctgtcgccca | gggtgagtgc | agtggtgtga | 3720 |
| | tcacaggtca | gtgtaacttt | taccttctgg | cctgagccgt | cctctcacct | cagcctcctg | 3780 |
| 5 | agtagctgga | actgcagaca | cgcaccgcta | cacctggcta | atttttaaat | tttttctgga | 3840 |
| | gacagggctc | tgtgtgtgtg | cccaggtctg | tctcaaactc | ttggactcaa | gggatccatc | 3900 |
| | tacctcggct | tcccaaagtg | ctgaattaca | ggcatgagcc | accatgtctg | gcctaatttt | 3960 |
| | caacactttt | atattcttat | agtgtgggta | tgtcctgtta | acagcatgta | gggtgaatttc | 4020 |
| 10 | caatccagtc | tgacagtcgt | tgtttaactg | gataacctga | tttattttca | tttttttgtc | 4080 |
| | actagagacc | cgcctgggtg | actctgattc | tccacttgcc | tgttgcatgt | cctcgttccc | 4140 |
| | ttgtttctca | ccacctcttg | ggttgccatg | tgcgtttcct | gccgagtgtg | tgttgatcct | 4200 |
| | ctcgttgcc | cctggctact | gggcatttgc | ttttatttct | ctttgcttag | tgttaccctc | 4260 |
| | tgatcttttt | attgtcgttg | tttgcctttg | tttattgaga | cagtctcact | ctgtcaccca | 4320 |
| | ggctggagtg | taatggcaca | atctcggctc | actgcaacct | ctgcctcctc | ggttcaagca | 4380 |
| 15 | gttctcattc | ctcaacctca | tgagttagctg | ggattacagg | cgcccaccac | cacgcctggc | 4440 |
| | taatttttgt | atttttagta | gagataggct | ttcaccatgt | tggccaggct | ggctctcaaa | 4500 |
| | tcctgacact | aagtgatctg | cccgcttgg | cctccacag | tgctgggatt | acaggtgcaa | 4560 |
| | gccaccgtgc | ccggcatacc | ttgatctttt | aaaatgaagt | ctgaaacatt | gctacccttg | 4620 |
| 20 | tcttgagcaa | taagaccctt | agtgtatttt | agctctggcc | acccccagc | ctgtgtgctg | 4680 |
| | ttttccctgc | tgacttagtt | ctatctcagg | catcttgaca | ccccacaaag | ctaagcatta | 4740 |
| | ttaatatgtt | tttccgtgtt | gagtgtttct | gtagctttgc | ccccgccctg | cttttctctc | 4800 |
| | tttgttcccc | gtctgtcttc | tgtctcaggc | ccgcctctg | gggtccccct | ccttgtctct | 4860 |
| | tgcgtgggtc | ttctgtcttg | ttattgcttg | taaaccocag | ctttacctgt | gctggcctcc | 4920 |
| 25 | atggcatcta | gcgacgtccg | gggacctctg | cttatgatgc | acagatgaag | atgtggagac | 4980 |
| | tcacagaggag | ggcgggtcatc | ttggcccgtg | agtgtctgga | gcaccacgtg | gccagcgttc | 5040 |
| | cttagccagt | gagtgcagc | aacgtccgct | cggcctgggt | tcagcctgga | aaacccagc | 5100 |
| | catgtcgggg | tctgggtggc | ccgcgggtgc | gagtttgaaa | tcgcgcaaac | ctgcgggtgtg | 5160 |
| | gcgccagctc | tgacgggtgt | gcctggcggtg | ggagtgtctg | cttcctccct | tctgcttggg | 5220 |
| 30 | aaccaggaca | aaggatgagg | ctccgagccg | ttgtcgccca | acaggagcat | gacgtgagcc | 5280 |
| | atgtggataa | ttttaaaatt | tctaggctgg | gcgcgggtgg | tcacgcctgt | aatcccagca | 5340 |
| | ctttgggagg | ccaaggcggg | tggatcacga | ggtcaggagg | tcgagaccat | cctggccaac | 5400 |
| | atgatgaaac | cccattctgt | ctaaaaacac | aaaaattagc | tgggcgtggg | ggcgggtgcc | 5460 |
| | tgtaatccca | gctactcggg | aggctgaggg | aggagaattg | cttgaacctg | ggagttggaa | 5520 |
| 35 | gttgacagtga | gccgacattg | caccactgca | ctccagcctg | gcaacacagc | gagactctgt | 5580 |
| | ctcaaaaaaa | aaaaaaaaaa | aaaaaaaaaa | aattctagta | gccacattaa | aaaagtaaaa | 5640 |
| | aagaaaagggt | gaaattaatg | taataataga | ttttactgaa | gccagcatg | tccacacctc | 5700 |
| | atcatttttag | gggtgtattg | gtgggagcat | cactcacagg | acatttgaca | ttttttgagc | 5760 |
| | tttgtctcgc | ggatcccggt | tgtaggtccc | gtgcgtggcc | atctcggcct | ggacctgctg | 5820 |
| 40 | ggcttcccat | ggccatggct | gttgtaccag | atgggtcagg | tccgggatga | ggtcgccagg | 5880 |
| | ccctcagtg | gctggatgtg | cagtgtccgg | atgggtgcag | tctgggatga | ggtcgccagg | 5940 |
| | ccctgctgtg | agctggatgt | gtggtgtctg | gatggtgcag | gtcaggggtg | aggtctccag | 6000 |
| | gccctcgggt | agctggaggt | atggagtcgg | gatgatgcag | gtccgggggtg | aggtcgccag | 6060 |
| | ggcctcgtgt | gagctggatg | tgtggtgtct | ggatgggtgca | ggtcaggggtg | gaggtctcca | 6120 |
| 45 | ggccctcggg | aagctggagg | tatggagtc | ggatgatgca | ggtcgggggtg | gaggtcgcca | 6180 |
| | ggccctgctg | tgagctggat | gtgtggtgtc | tggatgggtg | aggtctgggg | tgaggtcacc | 6240 |
| | aggccctcgg | gtgagctggg | tgtgcgggtg | ctggatgggtg | caggtctgga | gtgaggtcgc | 6300 |
| | cagacgggtg | cagaccatgc | ggtagctggg | atatgcgggtg | tccggatggg | gcaggtctgg | 6360 |
| | ggtagaggtg | ccaggccctg | ctgtgagttg | gatgtggggg | gtccggatgc | tgacaggtccg | 6420 |
| 50 | gtgtgaggtg | accaggccct | gctgtgagct | ggatgtgtgg | tgtctggatg | gtgcaggtct | 6480 |
| | gggggtgaagg | tcgccaggcc | cctgcttctg | agctggatgt | gtgggtgtctg | gatggtgcag | 6540 |
| | gtctggagtg | aggtcgccag | gccctcgggtg | agctggatgt | gcagtgtcca | gatggtgcag | 6600 |
| | gtccgggggtg | aggtcgccag | accctcgggtg | gagctggatg | tcgggtgtct | ggatgggtgca | 6660 |
| | ggctctggagt | gaggtcgcca | ggccctcggg | gagctggatg | tatggagtcc | ggatgggtgcc | 6720 |
| 55 | ggtcgggggtg | gaggtcgcca | gaccctgctg | tgagctggat | gtgcgggtgtc | tggatgggtac | 6780 |
| | aggtctggag | tgaggtcgcc | agaccctgct | gtgagctgga | tatgcgggtgt | ccggatgggtg | 6840 |
| | caggtcaggg | gtgaggtctc | caggccctcg | gtgagctgga | ggtagtgagt | ccggatgagt | 6900 |
| | caggtcgggg | gtgaggtcgc | caggccctgc | tgtgaactgg | atgtgcggcg | tctggatggg | 6960 |
| | gcaggtctgg | gggtgtgtcg | ccaggccctc | ggtagagctg | aggtatggag | tccggatgat | 7020 |
| 60 | gcaggtccgg | ggtaggtcg | ccaggccctg | ctgtgagctg | gatgtgcggc | gtctggatgg | 7080 |
| | tgcaggtctg | gggtgtgtgc | gccaggccct | cggtagctg | gaggtatgga | gtccggatga | 7140 |
| | tgcaggtccg | gggtgaggtt | gccaggccct | gctgtgagct | ggatgtgctg | tatccggatg | 7200 |
| | gtgcagtcgg | gggtgaggtc | gccaggccct | gctgtgagct | ggatgtgctg | tatccggatg | 7260 |
| | gtgcaggtct | gggtgaggt | caccaggccc | tgcgggtgagc | tgggtgtgag | gtgtccgggt | 7320 |
| 65 | gtgcaggtc | cgggtgaggt | tcggcgagcc | ctcgggtgagc | tggatgtgag | gtgtcccggt | 7380 |
| | gtccggatgg | tgacaggtcca | gggtgaggtc | gctaggccct | tgggtgggctg | gatgtgccgt | 7440 |

5 gtcaggatgg tgcaggctctg ggggtgaggtc gccaggccctt tggtagagctg gatgtgagg 7500
 gtctgcatgg tgcaggctctg ggggtgaggtc gccaggccctt tggtagagctg gatgtgagg 7560
 gtccggatgg tgcaggctccg gcgtgaggtc gccaggccctt gctgtgagct ggatgtgagg 7620
 tgtctggatg gtgcaggctcc ggggtgaggt agccaaggcc ttcggtagagc tggatgtggg 7680
 gtgtccggat ggtgcaggctc cgggtgaggt tgcaggccctt ctgaggcttag ctggatatgc 7740
 ggtgtccgga tgggtgcagggt ccgggtgaggt gtcaccaggc cctgaggctta gctggatgtg 7800
 cgggtgtctgg atggtgcagg tccgggtgga ggtcgccagg ccctgctgtg agctggatgt 7860
 gctgtatccg gatggtgcag gtccgggtg aggtcgccag gccctgcagt gagctggatg 7920
 10 tctgtatcc ggatggtgca ggtctggcgt gaggctgcca ggccctgagg ttagctggat 7980
 atgagggtgc ggatggtgca ggtccgggtg gaggctacca ggccctgagg ttagctggat 8040
 gtgagggtgc cggatggtgc aggtctgggg tgaggctgcc aggccctgct gtgagctgga 8100
 aagcgtgtat ccggatggtg caggctccggg gtgaggctgc caggccctgc ggtgagctga 8160
 atgtgctgta tccggatggt gcaggctctg cgtgaggctc ccaggccctg cggtagagctg 8220
 gatgtgcagt gtacggatgg tgcaggctcc gggtagaggtc gccaggccctt ggggtgggct 8280
 15 gtatgtgtgt tgtctggatg gtgcaggctcc ggggtgaggt cgccaggccc tgcggtagagc 8340
 tggatgtgtg gtgtctggat gctgcaggctc cgggtgaggt tgcaggccctt ctcggtagagc 8400
 tggatatgag gtgtcccggt gtccgaatgg tgcaggctcca gggtagaggtc gccaggccctt 8460
 tggtagagct gatgtgagg gtccggatgg tgcaggctctg gggtagaggtc gccaggccctt 8520
 20 tggtagagct gatgtgagg gtcaggatgg tgcaggctcc gggtagaggtc accaggccctt 8580
 cggtagatct gatgtggcat gtccttctcg ttttaag 8616

 <210> 6
 <211> 2089
 <212> DNA
 25 <213> Homo sapiens

 <400> 6
 gtactgtatc cccacgccag gcctctgctt ctggaagtc tggaaacacca gcccgccctc 60
 30 agcatgcgcc tgtctccact tgctctgctt tccctggctg tgcagctctg ggctgggagc 120
 cagggggcccc gtccacaggcc tgggtccaagt ggattctgtg caaggctctg actgcctgga 180
 gtccacgttc tcttacttgt aaaatcagga gtttgtgcca agtggctctt aggggttgta 240
 aagcagaagg gatttaaatt agatggaac actaccacta gcctccttgc ctttccctgg 300
 gatgtgggtc tgattctctc tctctttttt ttttctttt tgagatggag tctcactctg 360
 35 ttgcccaggc tggagtgcag tggcataatc ttggctcact gcaacctcca cctcctgggt 420
 ttaagcgatt caccagcctc agcctcctaa gttagctgga ttacaggcac ctgccaccac 480
 gcctggcctaa tttttgtact ttttaggagag acgggggttc accatgttgg ccaggctggt 540
 ctggaactca tgacctcagg tgatccaccc accttggcct cccaaagtgc tgggtttaca 600
 ggctaagcca ccgtgccag ccccgatctc tcttttaatt catgctgttc tgatgaatc 660
 40 ttcaatctat tggatttagg tcatgagagg ataaatccc acccacttgg cgactcactg 720
 cagggagcac ctgtgcaggg agcacctggg gataggagag ttccaccatg agctaacttc 780
 taggtggctg catttgaatg gctgtgagat tttgtctgca atgttcggct gatgagagt 840
 tgagattgtg acagattcaa gctggatttg catcagttag ggacgggagc gctggtctg 900
 gagatgccag cctggctgag cccaggccat ggtattagct tctccgtgtc ccgcccaggc 960
 45 tgactgtgga gggctttagt cagaagatca gggcttcccc agctcccctg cacactcgag 1020
 tccctggggg gccttgtgac acccatgcc ccaaatacag atgtctgcag agggagctgg 1080
 cagcagacct cgtcagaggt aacacagcct ctgggctggg gaccccgagc tgggtgctgg 1140
 gccatttctt tgcatctggg ggagggtcag ggccttccct gtgggaacaa gttataacac 1200
 aatgcacctt acttagactt tacacgtatt taatgggtgt cgacccaaca tggctatttg 1260
 50 accagtattt tggaaagaat ttaattgggg tgaccggaag gagcagacag acgtggtggt 1320
 ccccaagatg ctcttgttca ctactgggac tgttgttctg cctggggggc cttggaggcc 1380
 cctcctcccc ggacagggtg ccgtgccttt tctactctgc tgggcctgag gctgaggtc 1440
 agggaccag ctccggagca ccccgggccc cagtgtccac ggagtgcag gctgtcagcc 1500
 acagatcccc aggtccaggt gtggccgctc cagccccctg gccccatgg gtggttttg 1560
 55 gggaaaaggc caagggcaga ggtgtcagga gactgggtgg ctcatgagag ctgattctgc 1620
 tcttggctg agctgcccct agcagcctct cccgcccctc ccatctgaag ggatgtggt 1680
 ctttctacct ggggtcctg cctggggcca gccttgggt accccagtg ctgtaccaga 1740
 gggacaggca tctgtgtggt aggggcatgg gttcacgtgg cccagatgc agcctgggac 1800
 caggctcccc ggtgctgatg gtgggacagt caccctgggg gttgaccgcc ggactgggag 1860
 60 tccccaggtg tgactatagg accaggtgtc cagggtgcc gcaagtagag gggctctcag 1920
 aggcgtctgg ctggcatggg tggacgtggc cccgggcatg gccttcagcg tgtgtgccg 1980
 tgggtgccct gagccctcac tgagtcgggt ggggcttgtg gcttcccgtg agcttcccc 2040
 tagtctgttg tctggtgag caagcctcct gaggggctct ctattgcag 2089

10 / 18

<210> 7
<211> 687
<212> DNA
<213> Homo sapiens

5

<400> 7
gtggctgtgc tttggtttaa ctctcttttt aaacagaagt gcgtttgagc cccacatttg 60
gtatcagctt agatgaaggg cccggaggag gggccacggg acacagccag ggccatggca 120
cggcgccaac ccatttgtgc gcacagttag gtggccgagg tgccgggtgcc tccagaaaag 180
10 cagcgtgggg gtgtaggggg agctcctggg gcagggacag gctctgagga ccacaagaag 240
cagccggggc agggcctgga tgcagcacgg cccgaggtcc tggatccgtg tcctgctgtg 300
gtgcgcagcc tccgtgctgt tccgtttacg gggcccgggg accaggccac gactgccagg 360
agccqaccgg gctctgagga tcctggacct tgccccacgg ctctctgcacc ccacctctgt 420
ggctgcgggt gctgcgggtga ccccgctatc tggagagagt gtggggtgag gtggacagag 480
15 gtgtggcatg aggatcccg gtgcaacaca catgcggcca ggaacccgtt tcaaacaggg 540
tctgaggaag ctgggagggg ttctaggtcc cgggtctggg tggctgggga cactggggag 600
gggctgcttc tcccctgggt ccctatggtg ggggtgggac ttggccggat ccactttcct 660
gactgtctcc catgctgtcc ccgccag 687

20

<210> 8
<211> 494
<212> DNA
<213> Homo sapiens

25

<400> 8
gtgggtgccc gggacccccg tgagcagccc tgctggacct tgggagtggc tgcttgattg 60
gcacctcatg ttgggtggag gaggtactcc tgggtggggc gcagggagtg caggtgacct 120
tgctactgtt gaggacacac ctggcaccta ggggtggaggc ctccagcctt tcctgcagca 180
catggggccc actgtgcacc ctgactgccc gggctcctat tcccaaggag ggtccactg 240
30 gattccagtt tccgtcagag aaggaaccgc aacggctcag ccaccaggcc ccggtgcctt 300
gcacccagct cctgagccag gggctcctg tcctgaggct cagagagggg acacagcccg 360
ccctgccctt ggggtctgga gtggtggggg tcagagagag agtgggggac accgccaggc 420
caggccctga gggcagaggt gatgtctgag tttctgcgtg gccactgtca gtctcctcgc 480
ctccactcac acag 494

35

<210> 9
<211> 865
<212> DNA
<213> Homo sapiens

40

<400> 9
gtaaggttca cgtgtgatag tcgtgtccag gatgtgtgtc tctgggatat gaatgtgtct 60
agaatgcagt cgtgtctgtg atgcgtttct gtggtggagg tacttccatg atttacacat 120
ctgtgatatg cgtgtgtggc acgtgtgtgt cgtggtgcat gtatctgttg cgtgcatatt 180
45 tgtggtgtgt gtgtgtgtgg cacgtgtgtg tccatgggtg gtgtgcctgt ggtgtgcatt 240
tgtgtgtgtc tgtgacacgt gcatgttcat gctgtgtgct gcatgtctgt gatgtgccta 300
tttgtggtgt gtgtgtgcat gtgtccgtga catatgcgtg tctatggcat ggggtgtgtg 360
ggcccccttg ccttactcct tctcctccca ggcattggtc gcaccattgt cctcacgctc 420
tcgggtgctg gtttggggag ctccacattc agggctcctc cttctagcat ggggtgcccct 480
50 gtctgtgtac agggctgggc cttggagact gtaagccagg tttagagagga gagtagggat 540
gctggtggtg ccttctctga cccctggcac cccagggacc ccagtctggc ctatgccggc 600
tccatgagat ataggaaggc tgattcaggc ctgcgtcccc gggacacact cctccagag 660
cggccggggg ccttggggct cggcaggggt gaaaggggccc ctgggcttgg gttccacccc 720
agtgtgtcat agcacgctgg aggggtgaag cctcaaagtc gtgccaggcc ggggtgcaga 780
55 ggtgaagaag tatccctgga gcttcggtct ggggagaggc acatgtggaa acccacaagg 840
acctctttct ctgacttctt gagct 865

60

<210> 10
<211> 3782
<212> DNA
<213> Homo sapiens

<400> 10
tgtgggattg gttttcatgt gtgggatagg tgggatctg tgggattggt ttttatgagt 60
ggggtaaacac agagttcaag gcgagcttct tcctgtagt gggctctgag gtgctccaac 120

agctttattg aggagaccat atcttctctt gaactatggt cgggtttata gtaagtcagg 180
 ggtgtggagg cctcccctgg gctcccctgt ctgtttcttc cactctgggg tcgtgtgggt 240
 cctgctgtgg tgtgtggccg gtgggcaggg cttccaggcc tccttgtgtt cattggcctg 300
 gatgtggccc tggctacgct ccgtccttgg aattcccctg cgagtggag gctttcttct 360
 5 tttctttttt tctttctttt tttttttttt tgataacaga gtctcgctct tttttgccc 420
 ggctggagtg gtttggcggt atcttggctc actgcaacct gtgcttcctg agttcaagca 480
 attctcttgc ctcagcctcc caagtagctg gaattatagg cgtccaccac catgctgact 540
 aatttttcta atttttagtag agacgagggt tctccatgtt ggccaggctg gtctcgaact 600
 cctgacctca ggtgatcctc ccacctcggc ctcccaaagt gctgggatga cagggtgtga 660
 10 ccgcccgcgc cggccgagac tcgcttccct cagcttccgt gagatctgca gcgatagctg 720
 cctgcagcct tgggtgctgac aacctccgtt ttccttctcc aggtctcgct aggggtcttt 780
 ccatttcag actctcttca cagaagagtt tcacgtgtgc tgatttcccg gctgttctcc 840
 gcgtaatgg tgtctgtgtt ttatcgatgg cctccttcca tttcctttag gctttgttta 900
 ttgtgtttt tccggctcct tgaaggaaaa gtctcgatta tggatgtttg aactttcttt 960
 15 tctaaacaag catctgaagt tggcgttttc cctctaaagc agggatcccg agggccctgg 1020
 ctgtggagtg gcaccgggtc ggggcctgtt aggaaccggc cgcacagcgg gaggctaggt 1080
 ggggtgtggg gagccagcgt tcccgcctga gccccgccc tctcagatca gcagtggcat 1140
 gcggtgtcca gaggcgcaca caccctactg agaactgtgc gtgagagggg tctagattct 1200
 gtgctcctta tgggaatcta atgctgtgat atctgaggtg gaaccgtttg ctcccaaac 1260
 20 catcccttc cccactgtct tctgtgggaa aaatcgctct ccacgaaacc agtccctggg 1320
 accacaatgg ttggggaccg tgtgtctaaag acctgcttca gcagcctctc gtcagtgttg 1380
 atatatggc ttttctgtgt tgagtcacaga ataattacgg atttctgtga tgctttccgc 1440
 cgacctcaga cccatgggct atttgtgggc gtgttgccct ctcctgggtt gggaaggggtg 1500
 caggcccat gtaccttccct gttactgcct tccagggttg ttctcagggt tgaatcgta 1560
 25 tcgatgtgtt tttagccac ggccctgccc ccagctcctg ggggctgggg aacatgctga 1620
 agcacagagt caccgtgcgc gtcttttgat gcctcacaag ctcgagccct cctgtgtccg 1680
 tgttagtgtg tgtcacgtgc ctgctcacat cctgtcttgg ggacgcaggg gcttagcagg 1740
 tcccgtagta aatgacaagc gtcttggggg agtctgcaga ataggagggt ggggtgcccg 1800
 tctctctccc gcgtcttcag actcttctcc tgctgtgtct gtggctgcac ctgcatccct 1860
 30 gcaatccctc cagcactggg ctggagaggc ccgggagctc gactgccact tgtgccact 1920
 gactgtggat ggcagtcggg caggggggtc tgatgtgtgt gactgtgga tggcggttg 1980
 tcacaggggt ctgatgtgtg gtgactgtgt atggcggtcg tggggtctga tgtgtgact 2040
 gtggatggcg gtcgtggggg ctgatgtgtg gtgactgtgt atggcggtcg tggggtctga 2100
 tttggtgact tgggatggcg gtcgtggggg ctgatgtgtg gactgtggat gggggtctg 2160
 35 ggggtctgat tgggtactgt ggatggcagt cgtggggtct gatgtgtgt gactgtggat 2220
 ggcggctctg ggggtctgat tgggtactgt ggatggcagt cgtggggtct gatgtgtgt 2280
 gactgtggat ggcggctctg ggggtctgat tgtgtgtact gtggatggcg gtcgtgggg 2340
 ctgatgtgtg gtgactgtgt atggcggtcg tggggtctga tgtgtgtga ctgtggatg 2400
 cggctctggg gtcgtgatgt gtgactgtgt atggcggtcg tggggtctga tgtgtgtga 2460
 40 ctgtggatgg tgatcggta caggggtctg atgtgtgtgt actgtggatg gcggctctg 2520
 ggtctgatgt tgggtgactg tggatgggtga tccgtcacag ggggtctgat tgtgtgact 2580
 gtggatggcg gtcgtggggg ctgatgtgtg gtgactgtgt atggcggttg gtcccgggg 2640
 tctgatgtgt ggtgactgtg gatggcgatc ggtcacaggg gtctgatgtg tgggtactgt 2700
 ggatggcggt cgtgggggtc gatgtgtgt gactgtggat ggcgggtctg ggggtctgat 2760
 45 tgtgtgtact gtggatggcg gtcgtggggg ctgatgtgt gactgtggat ggcgggtctg 2820
 ggggtctgat tgggtactgt ggatggcggt cgtggggtct gatgtgtgt gactgtggat 2880
 ggcgggtgtt cccgggggtc tgatgtgtgt tgactgtgga tggcggtctg ggggtctgat 2940
 gtggtgactg tggatggcag tctgtgggtc tgatgtgtgt tgactgtgga tggcggtctg 3000
 ggggtctgat gtgtgtgtgac tgtggatggc ggtcgtgggg tctgatgtgt ggtgactgtg 3060
 50 gatggcggtc gtggggtctg atgtgtgtgt actgtggatg gcggtctgtg ggtctgatgt 3120
 ggtgactgtg gatggcggtc gtggggtctg atgtgtgtgt actgtggatg gtgatcgtg 3180
 acaggggtct gatgtgtgt gactgtggat ggcgggtctg ggggtctgat tgtgtgact 3240
 gtggatggcg gtcgtggggg ctgatgtgtg gactgtggat ggcgggtctg ggggtctgat 3300
 tgtgtgact gtggatggcg gtcgtagggt ctgatgtgt gtgactgtgt atggcagtcg 3360
 55 gtcacagggg tctgatgtgt ggtgactgtg gatggcggtc gtggggtctg atgtgtgtg 3420
 actgtggatg ggcgtctgtg ggtctgatgt gtgggtactg tggatggcg tctgtgggtc 3480
 tgatgtgtgt tgactgtgga tggcggtcgt ggggtctgat gtgggtactg tggatgggtg 3540
 tccgtcacag ggggtctgat tgtgttagct gcagggtggg tcccagggtg gtctgtagct 3600
 actttgcgtc ctcggccccc cggcccccgt ttcccaaac gaagcttccc aggcgtctc 3660
 60 tgggcttcat cccgccatcg ggcttggcgg caggtccaca cgtcctgac ggaagaaaca 3720
 agtggccagc tctggccggg gcaggccaca tttgtggctc atgcccctc ctctgccggc 3780
 ag 3782

12 / 18

<210> 11
<211> 980
<212> DNA
<213> Homo sapiens

5

<400> 11
gtctgggcac tgccttcgag gggtgggcac ggactcccag cagtgggtcc tcccctgggc 60
aatcactggg ctcatgaccg gacagactgt tggccctggg gggcagtggg gggaaatgagc 120
tgtgatgggg gcatgatgag ctgtgtgcct tggcgaaatc tgagctgggc catgccaggc 180
10 tgcgacagct gctgcattca ggcacctgct cacgtttgac tgcgcgccct ctctccagtt 240
ccgcagtgcc tttgttcattg atttgctaaa tgtcttctct gccagttttg atcttgaggc 300
caaaggaaag gtgtccccct cctttaggag ggcaggccat gtttgagccg tgtcttgccc 360
agctggcccc tcagtgtctg gtctgagccc aaaggaaacg tgtccccctt cttaggagga 420
cgggcctgtg ttgagccacg ccccgctgag cgggcctctc agtgcgtggg ctgtccacgt 480
15 ggccctgtgg ccctttgcag atgtggtctg tccacgtggc cctgtggctc tttgcagatg 540
cctgttagca cttgctcggc tctaggggac agtgcgtgtc accgcatgag gctcagagac 600
ctctgggcga atttccttgg ctcccagggt ggggggtggg gtggcctggg ctgtgggac 660
ccagaccctg tgcgggcgag ctgggcagca actcctggat cacatatgcc atccgggcca 720
cgggtggcgtg tgtgggtgtg agcccagctg gaccacaggt tggccagag gagacgttct 780
20 gtgtcacaca ctctgcctaa gcccattgtg gtctgcagag actcggcccg gccagccac 840
gatggccctg cattccagcc cagccccgca cttcatcaca aacactgacc ccaaaaggga 900
cggaggggtct tggccacgtg gtccctgctg tctcagcacc caccggctca ctcccatgtg 960
tctcccgtct gctttcgag 980

25

<210> 12
<211> 2485
<212> DNA
<213> Homo sapiens

30

<400> 12
gtgagtcagg tggccagggt ccattgcctt gcgggtgggt gggcgggctg gcagggtctc 60
tgctcacctc tctcctgccc ctccccact gnccttctgc ccggggccac cagagtctcc 120
ttttctggcc cccgccccct ccggctctct ggtgcaggc tcccgaggcc ccggaacat 180
ggctcggctt gcggcagccg gagcggagca ggtgccacac gaggcctgga aatggcaagc 240
35 ggggtgtgga gttgtcctct cgtggaggac gaggggcggg ggggtgtgtc gggtcagggt 300
tgcgcccagc gtttgagcct gcagctgtc agctccaaat tactactgac gctggacacc 360
cggtctcac acgcttgtat ctctctctcc cgatacaaaa ggattttatc cgattctcat 420
tcctgtccct gtctgtgtgac ccccgcgagg gcgcggtctc ttctctctgt gactagattt 480
ccatctgga aagtgcgggg ttgaccgtgt agtttgctcc tctcgggggg cctgtggtg 540
40 ccatggggca ggcggcctgg gagagctgcc gtcacacagc cactgggtga gccacactca 600
cgggtggtaga gccacagtgc ctgggtgccac atcacgtcct ctggatttta agtaaaacca 660
cacacctccc ggcaggcatc tgcctgcgac cctgtgtgtg cctggggaga gtggtagcac 720
ggaggaaatt cgtgcacact caaggtcatc agcaaggta tccgcagtca ggtggaacgt 780
ggaggcctct ctctgggacg gtctccagcg gataaaggac tgtgcacagc ttcggaagct 840
45 ttattataaa aatataacta ttaattattg cattataagt aatcactaat ggtatcagca 900
attataatat ttattaaagt ataattagaa atattaaagta gtacacacgt tctgaaaaaa 960
cacaaattgc acatggcagc agagtgaatt ttggcggagg gacacgtgtg cacatgtgtg 1020
taagcggccc ccaggccac agaattcgtc gacaaagtca cctccccaga gaagccacca 1080
cgggcctcct tctgtgtcgt gaattttatt aagatggatc aagtcacgta ccgtccacgt 1140
50 gtggcagggc tttggggaat gtgaggtgat gactgcgtcc tcatgccctg acagacagga 1200
ggtgactgtg tctgtcctgt ccctaggaca cggacaggcc cgaagctcta gtccccatcg 1260
tggctccagt tggcctctga ataaaaacgt ctcaaaaacc tgtgccccca aaaaactaaga 1320
acagagagag tttcccatcc catgtgtcca caggggcgta tctgcttgcg ttgactcgct 1380
gggctggccg gactcctaga gttggtgcgt gtgcttctgt gcaaaaagtg cagtccctct 1440
55 gcccactact gtgatctctg caccagcaag gaaagcctct tttcttttct tctttttttt 1500
ttttttgaga cggaaacgtca ctgtgtctg cctgggcttg agtgcagtgg cgcatctca 1560
actcactgca acctccgect cccgggttcc agcattttct ctgcctcagc ctcccagca 1620
gctgagatta caggcaccca cccctgcgc ctggctaatt tttgtatttt tagtagagag 1680
gggttttttc catgttgccc aggtgtgtct cgaactcctg acctcagggt atccaccac 1740
60 ctccggcctcc caaagtgtg ggattacagg tgtgagccat caccgccagc cggaaagcct 1800
ctttttaagg tgaccaccta tagcgcttcc cgaaaataac aggtcttgtt tttgcagtag 1860
gctgcaagcg tctcttagca acaggagtgg cgtcctgtgg gctctgggga tggctgaggg 1920
tcgctgggca gccatgcctt ctgtgtgcac ctttaggttc caccgggcta ttctgtctc 1980
actgtttgtc tgaaaacgca cccttggcat ccttgttttg agagtttctg cttctcgttg 2040
65 gtcattgtga aactaggggc aaggttgtat ccgttggcgc gcagcggcta catgtagggg 2100

13 / 18

5 catgagtctt tcaccgtgga caaatcctt gaaaaaaaa aaaggagtcc ggtaagcat 2160
 tcattccggg tcaagtgtct ggttctgtga ataaactcta agatttaaga aaccttaatg 2220
 aaagaaaacc ttgatgattc agagcaagga tgtgggcaca cctgtggctg gatctgtttc 2280
 agccgcccc gtgcatggtg agagtgggga gcagggtattg tttgttcaga ggtctcatct 2340
 ggatgtttc tgaggtgttt gccggctgaa tggtagacgt gtcgtttgtg tgtatgaggt 2400
 tctgtgtctg tgtgtggctc ggtttgagtg tacgcatgtc cagcacatgc cctgcccgtc 2460
 tctcacctgt gtcttccgc cccag 2485

10 <210> 13
 <211> 1984
 <212> DNA
 <213> Homo sapiens

15 <400> 13
 gtgaggcctc ctcttcccca ggggggcttg ggtgggggtt gatttgcttt tgatgcattc 60
 agtgtaata ttcttggtgc tctggagacc atgactgtc tgtcttgagg aaccagacaa 120
 ggttgacgcc cttcttgggt atgaagccgc acgggagggg ttgcacagcc tgaggactgc 180
 gggctccacg caggctctgt ccagggcca tgtccagagg cctcagggt cagcaggcgg 240
 gagggccgct gccctgcatg atgagcatgt gaattcaaca ccgaggaagc acaccagctt 300
 20 ctgtcacgtc acccaggttc cgtaggggtc cttggggaga tggggctggt gcagcctgag 360
 gcccacatc tcccagcagg ccctcgacag gtggcctgga ctgggcgctt cttcagccca 420
 ttgcccattc cacttgcatg ggttctacac ccaaggacgc acacacctaa atatcgtgcc 480
 aacctaatgt ggttcaactc agctggcttt tattgacagc agttactttt ttttttttaa 540
 tactttaagt tctaggttac atgtgcacga cgtgcagggt agttacatat gtatacatgt 600
 25 gccatgttgg tgtgctgcac ccattaactc atcatttaca ttaggtatat ctccaatgc 660
 tatccctccc cactccccc atcccatgac aggccctggt gtgtgatgtt cccaccctg 720
 tgtccaagt tctcattgt tcagttccca cctgtgagtg agaacatgtg gtgtttggtt 780
 ttctttcctt gcaatagttt gctcagagtg atggtttcca gcttcgtcca tgtccctaca 840
 aaggacatga actcatcctt ttttatgact gcatagtatt ccgtggtgta tatgtgccac 900
 30 attttcttaa tccagtctat catcgatgga catttgggtt ggttgcaagt ctttctact 960
 gtgaatagtg ccgcaataaa catacgtgtg catgtgtctt tatagcagca tgatttataa 1020
 tcctttgggt atataccag taatgggatg gctgggtcaa atggtatttc tagttctaga 1080
 tccttgagga atcaccacac tgtcttcac aatggttgaa ctagtttaca ctcccaccaa 1140
 cagtgtaaaa gtgttctggt gctggagagg atgtggacag cagttatttt tttatgaaaa 1200
 35 tagtatcact gaacaagcag acagttagtg aaggatgctg caggaagcct gcaggccaca 1260
 cagccatttc tctcgaagac tccgggtttt tctgtgcat cttttgaaac tctagctcca 1320
 attatagcat gtacagtgga tcaaggttct tcttcattaa ggttcaagtt ctgagattgaa 1380
 ataagtttat gtaacagaaa caaaaatttc ttgtacacac aacttgctct gggatttgga 1440
 ggaaagtgtc ctcgagctgg cggcacactg gtcagccctc tgggacagga tacctctggc 1500
 40 ccatggtcat ggggcgctgg gcttggcct gaggtcaca cagtgcacca tgcccagctt 1560
 cctgtggata ggatctgggt ctcgatcat gctgaggacc acagctgcca tgctggtaaa 1620
 gggcaccacg tggctcagag gggcgaggt tcccagcccc agctttctta ccgtcttcag 1680
 ttatttttcc ctaagagtct gagaagtggg gccgcgctg atggccttcg ttcgtcttca 1740
 gctggcacag aattgcacaa gctgatggt aacactgagt acttataatg aatgaggaat 1800
 45 tgctgtagca gttactgta gagagctcgt ctgttgaaa gaaatttaag tttttcattt 1860
 aaccgctttg gagaatgtta cttattttat ggctgtgtaa attgtttgac attcagctcc 1920
 tcgtagacag atactacgta aaaagtgtaa agttaacctt gctgtgtatt ttcccttatt 1980
 ttag 1984

14 / 18

5 gggattgtcc aatgtggtcc ccctcaaggc cgccccacag agccggtggg cttgttttaa 660
 agtgcgattt gacgagggac gagaaacctt gaaagctgta aagggaaacc tcagaaaatg 720
 tggccgccag ggggtggttc aggtgctttg ctgggctgtg tttgtgaaaa ccattttgga 780
 cccgccctcc aagtccaccc tccaggtcca ccctccaggc ccgccctggg ctgggggtat 840
 gcctggcggt ccttgtgccc cagcccgagc cacagcaggc tgtgcacatt taaatccact 900
 aagattcact cgggggggagc ccaggtccca agcaactgag ggctcaggag tcctgaggct 960
 gctgagggga cagagcagac ggggaacgct gcttctgtgt ggcaagttcc tgaggggtgt 1020
 gggcagggag gtggctcaga gtgtatgttg ggggtcccacc gggggcagaa ctctgtctct 1080
 gatgagtcgg cagccatgta acaggaaggc gtggccacag ggagctggga atgcaccagg 1140
 10 ggagctgcgc agctggccga ggtcccaggc ccaggccaca ggaagggcag ggggacgccc 1200
 ggggccacag cagaggccgc aggaaggga ggggatgcc aggccagagc agaggctacc 1260
 gggcacaggc ggggtccctg agctgggtga gcgaggctca tgactcggcg agggaaacct 1320
 cttgacgtga agctgacgac tgggtgtgcc cagctcacag ccagccagg tcccgcgcct 1380
 gagcaggaac tcagaacctt cccctttgtc taaagcacag cagatgcctt cagggcactt 1440
 15 aggagaaaac agggaaagtc gttgagaaac gtcttaaaag aagggtggat ggtggcaatt 1500
 tcttgtccag attttagtct gccccggacc acagatgagt ctataacggg atttgtgtgt 1560
 tgccatgggg acacatgaga tggaccatca cagaggccac tggggtgca cctcccatct 1620
 gagtctggc tgtcccggt ccaggccagg ttcttgcatg ctcacctacc tgtctgccc 1680
 gggagacagg gaaagcaccc cgaagtctgg agcagggctg ggtccaggct cctcagagct 1740
 20 cctgccaggc ccagaccct gctccaaac accacttctc tggggtttcc caaagcattt 1800
 aacaaggggt tcaggttacc tcctgggtga cggccccga tcctggggct gacattgccc 1860
 ctctgcctta g 1871

25 <210> 15
 <211> 3801
 <212> DNA
 <213> Homo sapiens

30 <400> 15
 gtgagcgac ctggccggaa gtggagcctg tgccccgctg gggcaggtgc tgctgcaggc 60
 ccgttgctc caectctgct tccgtgtggg gcaggcgact gccaatccca aagggtcaga 120
 ggccacaggc tgcccctcgt cccatctggg gctgagcaga aatgcattt tctgtgggag 180
 tgaggggtgt cacaacggga gcagttttct gtgctatttt ggtaaaagga aatggtgcac 240
 cagacctggg tgactgagg tgtcttcaga aagcagtctg gatccgaacc caagacgccc 300
 35 gggccctgct gggcgtgagt ctctcaaacc cgaacacagg ggccctgctg ggcatgagtc 360
 cctctgaacc cgagaccctg gggccctgct gggcgtgagt ctctccgaac ccagagactt 420
 cagggccctt ttgggcgtga gtctctccgc tgtgagcccc acactccaag gctcatccac 480
 agtctacagg atgccatgag ttcatgatca cgtgtgacct atcaggggac agggccatgg 540
 tgtggggggg gtctctacaa aattctgggg tcttgtttcc ccagagcccc agagctcaag 600
 40 gcccctctc aggtcagac acaaatgaat tgaagatgga cacagatgca gaaatctgtg 660
 ctgtttcttt tatgaataaa aagtatcaac attccaggca gggcaagggt gctcacacct 720
 ataatcccag cactttggga ggccgagggt ggtggatcac ttgaggccag gattttgagg 780
 ccaacctaac caacatagtg aaattccatt tctacttaaa aaatacaaaa attagcctgg 840
 cctggtggca cagcctgta gtccccgcta tgccggaggc tgaggcagga gaatcattt 900
 45 aaccaggag gcagagggtg cagtgagccg agatcacacc actgcactcc agcctgggca 960
 acagagttag acttcatctt aaaaaaaaaa aaaaaagtat cagcattcca aaaccatagt 1020
 ggacagggtg ttttttattc tgccttcga taatatttac tgggtgctgtg ctgaggccg 1080
 gaactggggg tgccttctc tgaaggcac accttcatgg gaagagaaat aagtgtgaa 1140
 tgggtgttaa accagaggtt taaactgggg tcctgtcgtt ctgagttaac agtccagatc 1200

15 / 18

| | | | | | | | |
|----|--------------------|-------------|------------|-------------|-------------|------------|------|
| | gactggaagc | aaataagttg | tgtctttaca | gcatatacca | gagcagattc | taggtagaag | 2220 |
| | aggagacaca | tgcaaaacaac | accagcaaca | gaaataaaaac | aaaagactca | aaggggaagg | 2280 |
| | aggtgaacgt | tccttggttt | ggtgttgagg | aaggacacac | agggaggcgg | atgaaaccag | 2340 |
| 5 | tgaggcaacg | ggcattgctt | tacttgacga | gaaactcagc | ttgcctgagc | cacagtgaag | 2400 |
| | atggccattc | cctggagcgt | ttgtgcacgt | gatttattta | aggcgccctg | tgaggtcctg | 2460 |
| | cacattcatt | ctctcacttt | gttctcctaa | ccacctgaga | ggtagaggag | gaaaggctcc | 2520 |
| | aggggagcag | ccgcccttgg | tcacccagct | ggcaaagggc | atgcatgatt | gcagcctggc | 2580 |
| | ctcctgctcc | ggggcccttg | ctctgcccga | ggacccca | caagtcagac | ccataggctc | 2640 |
| 10 | aggggtgagc | ggagcccaag | gtcgtgttgg | ggatggctgt | gaaagaagaa | atggacgtct | 2700 |
| | gatgcacact | tggaaggtc | ctaccagcag | cgtcaaagaa | atgcatgtga | aactgacagc | 2760 |
| | gagacccatc | cctcaaagaa | acgcacgtga | aactgatggc | gagacctgtc | cccatccctc | 2820 |
| | atgctggctc | cttttctggg | cttgccaaga | gccagcatca | ggttgaggca | agctggaaag | 2880 |
| | acttttctgg | aaagcagctt | gtttgcatgg | aagtcctcac | aatgtcctgt | gtcttcccag | 2940 |
| 15 | taattccact | tctgaagtga | ccagacatta | tcacgggtct | tatttaccat | ttccagtgtt | 3000 |
| | ccaggcaggg | ggacttgcca | cagcaagtca | cgaacctgcc | caaatacagg | gctaaggaga | 3060 |
| | tattatgcat | cacaaaactt | gctctgccat | taaacatttt | tcaaagaatt | tttgaagaat | 3120 |
| | gtttaatggc | acaaaacgtt | tatttcaatg | tagcagtgtt | caaagctgga | tgtaaaagaa | 3180 |
| | cacacccag | gagcctgccc | tgaatgtcat | gtgtgttcat | ctttggacat | ggacatacat | 3240 |
| 20 | gggcagtggg | tggtgtgag | gccctggagg | acatcggtgg | gatgcctcca | tcctgcccct | 3300 |
| | ctggagacac | catgtgtgcc | acgtgcactc | actggagccc | tgtttagctg | gtgccacctg | 3360 |
| | gctcttccat | ccctgagatt | caaacacagt | gagattcccc | acgcccact | cagtgttctc | 3420 |
| | ccacaaaaaa | cctgagtcac | acctgtgttc | actcgaggga | cgcccgggag | ccagggctcc | 3480 |
| | acagtttatt | atgtgttttt | ggctgagtta | tgtgcagatc | tcacagggc | agatgatgag | 3540 |
| 25 | tgcacaaaca | cgccgtgccc | aggtttggat | acactcaaca | tcactagcca | ggtcctgggt | 3600 |
| | gagtttgggt | atgcagagtc | tgatggcat | gtagcatttg | gagtcctagg | agtgagcacc | 3660 |
| | cagccccctc | gggctgcagc | gcatgcccga | ggcaggacaa | ggaagcggga | ggaaggcagg | 3720 |
| | aggctctttg | gagcaagctt | tgcaggaggg | ggctgggtgt | ggggcaggca | cctgtgtctg | 3780 |
| | acattcccc | ctgtgtctca | g | | | | 3801 |
| 30 | <210> 16 | | | | | | |
| | <211> 880 | | | | | | |
| | <212> DNA | | | | | | |
| | <213> Homo sapiens | | | | | | |
| 35 | <400> 16 | | | | | | |
| | gtgagcaggc | tgatggtcag | cacagagttc | agagtccagg | aggtgtgtgc | gcaagtatgt | 60 |
| | gtgtgtgtgt | gtgcgcgcgt | gcctgcaagg | ctgatgggtga | ctggctgcac | gtaagagtgc | 120 |
| | acatgtacgc | atatacacgt | gagcacatac | atgtgtgcat | gtgtgtacat | gaaggcatgg | 180 |
| 40 | cagtgtgtgc | acaggtgtgc | aagggcacaa | gtgtgtgcac | atgcgaatgc | acacctgaca | 240 |
| | tgcattgtgtg | ttcgtgcaca | gtcgtgtggg | cattcacgtg | aggtgcatgc | gtgtgggtgt | 300 |
| | gcagtgtgag | tagcatgtgt | gcacataaca | tgtattgagg | ggtcctcgtg | ttcaccccgc | 360 |
| | taggtcctca | gcaccagtgc | cactccttac | aggatgagac | gggggtcccag | gccttgggtg | 420 |
| | gctgaggctc | tgaagctgca | gccctgaggg | cattgtccca | tctgggcac | cgctccact | 480 |
| 45 | ccctcctctg | tggtcttctg | tgtccactcc | ccctcctctg | tggtgattta | catccactcc | 540 |
| | actcctctc | tcctgtgggc | atccgcgtcc | actccccctc | tctgtgggca | tctgcgtcca | 600 |
| | ctccccctct | ctgtgggcat | ttgcgtccac | tccctcctct | ggttccttcc | tgtcttggcc | 660 |
| | gagcctcggg | ggcaggcaga | tgacacagag | tcttgactcg | cccagggtgg | ttcgcagctg | 720 |
| | ccgggtgagg | gccaggccgg | atttactctg | gaagagggat | agtttcttgt | caaaatgttc | 780 |
| 50 | ctctttcttg | ttccatctga | atggatgata | aagcaaaaag | taaaaactta | aaatcccaga | 840 |
| | gaggtttcta | ccgtttctca | ctctttcttg | gcgactctag | | | 880 |
| | <210> 17 | | | | | | |
| | <211> 3186 | | | | | | |
| | <212> DNA | | | | | | |
| 55 | <213> Homo sapiens | | | | | | |
| | <400> 17 | | | | | | |
| | gtgagccgcc | accaaggggt | gcaggccag | cctccaggga | ccctccgcgc | tctgctcacc | 60 |
| | tctgaccggg | ggcttcacct | tggaactcct | gggttttagg | ggcaaggaa | gtcttacgtt | 120 |
| 60 | ttcagtggtg | ctgctgcctg | tgacagttc | tgctcgcgtg | gctctgtgca | aagcacctgt | 180 |
| | tctccatctc | tggttagtgg | taggagccgg | tggtggccca | ggtgtcccca | ctgtgcctgt | 240 |
| | gcactggccg | tggtgactga | tggtggccat | cccagggcag | caggggcatg | gggtaaagag | 300 |
| | atgtttatgg | ggagtcttag | cagaggaggc | tggtgaaggtg | tctgaacagt | agatgggaga | 360 |
| | tcagatgccc | ggaggatttg | gggtctcagc | aaagagggcc | gaggtgggtg | caggtgaggg | 420 |
| 65 | tcgttgcccc | cacccccggg | aaggtgcagc | agagctgtgg | ctccccacac | agccccggca | 480 |

5 gcacctgtgc tctgggcatg gctgtgctcc tggaaacgttc cctgtccttg ctggtcaggg 540
 ggtgcccctg ccaagaatcg acaactttat cacagaggga agggccaatc tgtggaggcc 600
 acagggccag cttctgcctg gactcagggc aggtggtggc acaagcctcg gggctgtacc 660
 aaagggcagt cgggcaccac agggccgggc ctccacctca acaggcctcc cgagccactg 720
 ggagctgaat gccaggaggc cgaagccctc gcccctatgag ggctgagaag gactgtgagc 780
 atttgtgtta cccagggccg aggtgcgcg aattaccgtg cacacttgat gtgaaatgag 840
 gtcgtcgtct atcgtggaaa cccagcaagg gctcacggga gacttttcca ttacaaggtc 900
 gtaccatgaa aatgggtttt aaccagagt gctgcgcctt catgctctgg caggaggaggc 960
 10 agagccacag ctgcatgtta ccgcttttgc accagctcca gaggtctggg accaggctgt 1020
 ctccagttcca ggggtgcgtc ggctcagacc gccctcctct ctgctctctc tctctgcctc 1080
 aaatcttccc tctgttgcag ctccctgacg cgtgcctggg ccctcgtgca agctgcttga 1140
 ctcccttccg gaaacccctg ggggtgtgct gatacaggtg ccactgagga ctggagggtg 1200
 ctgacactgt ggttgacccc aggttcacg tggcgtgctt ggggcctcct tgggcatga 1260
 tgaggtcaga ggagttttcc cagggtgaaa ctctgggaa actcccaggg ccatgtgacc 1320
 15 tgccacctgc tctctccata ttacgtctag tctgtcctc atttcccac cagggtctct 1380
 agctccgagg agctcccga gagggcctg gctcaggga gggcggtga gtttcccac 1440
 ccatgtgggg acccttgggt agtcgcttga ttgggtagcc ctgaggaggc cgagatgca 1500
 tggggccacg gccgtttcca aacacagagt caggcacgtg gaaggcccag gaatcccct 1560
 ccctcgaggc aggagtggga gaacgagag ctgggcccgc atttcacggc agccaggctg 1620
 20 cagtgggcga ggcgtgtgtg gtccacgtg cgctgggggc ggggtctgat tcaaatccgc 1680
 tggggctcgg ccttctggc ccgtgctggc cgcgcctcca cacgggctg ggggtgacgc 1740
 cccgacctct agcagggtgc tatttctccc ttggaagag agcccctcac ccatgctagg 1800
 tgttccctc ctgggtcagg agcgtggccg tgtggcaacc ccgggacctt aggttattt 1860
 atttgtttaa aaacattctg ggcctggctt ccgttgttgc taaatgggga aaagacatcc 1920
 25 cacctcagca gacttactga gaggtgaaa ccggggtgct ggcttgactg gtgtgatctc 1980
 aggtcattcc agaagtggct caggaaagta gtgagaccag gtacatgggg ggctcaggca 2040
 gtgggtgaga tgaggtacac ggggggctca ggcagtgggt gagggccagg acatgggggg 2100
 ctcaggcact ggggtgagat aggtacacgg ggggctcagg cagagggtca gaccaggtag 2160
 acgggggctc tgatcacacg cacatatgag cacatgtgca catgtgctgt ttcatggtag 2220
 30 ccaggctctgt gcacacctgc cccaaagtcc caggaaagt agaggccaaa gatggaggct 2280
 gacagggctg gcgcgtgggc tcacacctgt agtcccagca ctttgggagg ccgaggcgag 2340
 aggatccctt gagcccagga gtttaagacc agcctgagca acatagtaga accccatctc 2400
 tatgaaaaa aaaaacaaa attagctgaa catggtgggt tgcgcctgta gttccaatac 2460
 35 ttgggaggct gaagtgggag gatcacttga gcccaggagg tggaaagtgc agtgagctga 2520
 gattgcacca ctgtactgca gcctgggtga cagagtga gcccctctca acaacaacaa 2580
 agaagactga caaatgcagt ttcttggaaa gaaacattta gtaggaaact aacctacaca 2640
 cagaagccaa gtcggtgtct cgtgtgctag gagatgagat gatgggtcct cacaccatca 2700
 cccagacccc aggggtttatg caccacaggg gcgggtggct cagaagggat gcgcaggacg 2760
 ttgatatacg atgacatcaa ggtgtcttga cgaagggcag gattcatgat aagtacctgc 2820
 40 ttgtacacaa ggaacaatgg ataaactgga aaccttagag gccttcccgc aacaggggct 2880
 aatcagaagc cagcatgggg ggctggcatc caggatggag ctgcttcagc ctccacatgc 2940
 gtgttcatac agatggtgca cagaaacgca gtgtacctgt gcacacacag acacgcagct 3000
 actcgcacac acaagcacac acacagacat gcattgcatgc atccgtgtgt gtgcacctgt 3060
 gcccattgag aaacccatgc atgtgcattc atgcacgcac acaggcaccg gtggggccat 3120
 45 gcccacaccc acgagcaccc tctgattagg aggcctttcc tctgacgctg tccgccatcc 3180
 tctcag 3186

<210> 18

<211> 781

<212> DNA

<213> Homo sapiens

50

<400> 18

55

gtatgtgcag gtgcctggcc tcagtggcag cagtgcctgc ctgctgggtg tagtgtgtca 60
 ggagactgag tgaatctggg cttaggaagt tcttaccctc ttctgcacga ggaagtgggt 120
 taaccaaac actgtcaggc tctgtctgcc gccctctcgt ggggtgagca gagcactga 180
 tggaggagac aggagctgtc tgggagctgc catccttccc accttgcctc gcctggggaa 240
 gcgctggggg gcctgggtctc tctgttttgc cccatgggtg gatttggggg gcctggcctc 300
 tctgttttgc cctgtgggtg gattgggctg tctcccgtcc atggcactta gggcccttgt 360
 60 gcaaaccag gccaaaggct taggaggag ccaggcccag gctaccccac cctctcagg 420
 agcagaggcc gcgtatcacc acgacagag cccgcgcctg cctctgcttc ccagtcaccg 480
 tctctgccc ctggacactt tgtccagcat caggagggtt tctgatccgt ctgaaattca 540
 agccatgtcg aacctgcggg cctgagctta acagcttcta cttctgttcc tttctgtgtt 600

17 / 18

gtggaaat ttt cacctggaga agccgaagaa aacatttctg tctgtactcc tgcggtgctt 660
gggtcgggac agccagagat ggagccaccc cgcagaccgt cgggtgtggg cagctttccg 720
gtgtctcctg ggaggggagc tgggctgggc ctgtgactcc tcagcctctg ttttccccc 780
g 781

5
<210> 19
<211> 536
<212> DNA
<213> Homo sapiens

10
<400> 19
gcaagtgtgg gtggaggcca gtgcggggccc cacctgcccc ggggtcatcc ttgaacgccc 60
tgtgtggggc gagcagcctc agatgctgct gaagtgcaga cggccccggg cctgaccctg 120
ggggccttga gccacgttgg cagccctatg tgattaaacg ctggtgtccc caggccacgg 180
15 agcctggcag ggtccccaac ttcttgaacc cctgttccc atctcagggg cgatggctcc 240
ccacgttgg gagccttctg acccctgacc tgtgtcctct cacagcctct tccctggctg 300
ctgcccctgag ctctctgggt cctgagcaag ttctctcccc gccccgcgcg tccagcgtca 360
ctgggctgcc tgtctgctcg ccccgggtga ggggtgtctg tcccttact gaggttccc 420
ccagccaggg ccacgaggtg caggccctgc ctgcccggcc acccacacgt cctaggaggg 480
20 ttggaggatg ccacctctg cctcttctg aacggagtct gatattggcc ccgcag 536

<210> 20
<211> 3179
<212> DNA
<213> Homo sapiens

25
<400> 20
atctcatgtt tgaatcctaa tgtgcactgc atagacacca ctgtatgcaa ttacagaagc 60
ctgtgagtga acgggggtgt ggtcagtgcg ggcccatggc ctggctgtgc atttacggaa 120
30 gtctatgagt gaatggggtt tgggtcagtg cgggcccatt gcctggctgg gcctgggagg 180
tttctgatgc tgtgaggcag gaggggaagg agggtagggg atagacagtg ggagccccc 240
ccctggaaga cataacagta agtccaggcc cgaaggcgag cagggatgct gggggcccag 300
cttgggcggc ggggatgatg gagggcctgg ccagggtggc agggatgatg gggggcccag 360
ctgggggtggc aggggtgatg gggggggctg gtctgggtgg cggggaagat ggggaagcct 420
35 ggctgggccc cctctcccc tgcctccac ctgcagccgt ggatccggat gtgcttccct 480
gggtgcacatc ctctgggcca tcagctttca tggaggtggg gggcaggggc atgacacat 540
cctgtataaa atccaggatt cctctctctg aacgcccaca ctcaggttga aagtcacatt 600
ccgctctctg ccattctctt aagagtagac caggattctg atctctgaag ggtgggtagg 660
gtggggcagt ggagggtgtg gacacaggag gcttcagggt ggggctgggt atgctctctc 720
40 atcctcttat catctccag tctcatctct catctctta tcctctccca gtctcatctg 780
tcttctctct atctccagt ctcatctgtc atctcttac catctccag tctcatctct 840
tctctctta tctctagt ccatccagac ttacctcca gggcggtgct caggctcgca 900
gtggagctgg acatacgtcc ttcctcaggc agaaggaact ggaaggattg cagagaacag 960
gagggcgggc tcagagggac gcagtcttgg ggtgaagaaa cagccctcc tcagaagtgc 1020
45 gcttgggcca cagaaaccg agggccctgc gtgagtggct ccagagcctt ccagcaggtc 1080
ccttggtggg ccttatggta tggccgggtc ctactgagt caccttggac agggcttctg 1140
gtttgagtgc agcccggagc tgcctgggtg cgggggtggg gcttatggcc actggatatg 1200
gogtcattta ttgtgtctgc ttcagagaat gtctgagtga ccgagcctaa tgtgtatggt 1260
gggcccaggt ccacagactg tgtcgtaaat gcactctggt gcctggagcc cccgtatagg 1320
50 agctgtgagg aaggaggggc tcttggcagc cggcctgggg gcgcctttgc cctgcaaac 1380
ggaaggaggc ggccccgggc gccgtgggag gacgacctca agtgagaggt tggacagaac 1440
agggcgggg cttcccagga gcagaggccg ctgctcaggc acacctgggt ttgaatcaca 1500
gaccaacagg tcaggccatt gttcagctat ccacttctta caaagctcca gattcctgtt 1560
tctccgggtg ttttttgttg aaattttact caggattact tatatttttt gctaaagtat 1620
55 tagaccctta aaaaaggtat ttgctttgat atggcttaac tcactaagca cctactttat 1680
ttgtctgttt ttatttatta ttattattat tattagagat ggtgtctact ctgtcacc 1740
ggttgttagt gcagtggcac agtcatggct cgctgtagcc gcaaaccccc aggtc 1800
gacccctcgg cctcagcttc ccagagtgtg gggattacag gtgtgagcca ctgcccctgc 1860
ctggcacttt taaaaaccac tatgtaaggt caggtccagt ggcttcaca cctgtcatcc 1920
60 cagtattttg ggaagccgag gcagaaggat tgtctgaggc caggagtgtg agaccagcat 1980
gggtaacata gggagacccc atctctacaa aaaatgcaaa aagttatccg ggcgtgggg 2040
ccagcatctg tagtcccagc tgcctgggag gctgagtggg aggatcgctt gagcccgga 2100
ggcctatggc gcagttagct gtgattgtac catcgactc cagcctgggc aacagagtga 2160
gaccctgtct caaaaaaaaa aaaaaaaaaa gaaggagaag gagaagagaa gaagaaggaa 2220
65 gaaggaaaga gaagaagaag gaagaaggaa gaaagaaggaa gaaggaggcc tgctagggtc 2280

18 / 18

5 taggtagact gtcaaattctc agagcaaaat gaaaataaca aagttttaaa gggaaagaaa 2340
aaccacagct ctttggactt ccttaggcct gaacttcac tcaagcagct tccttcacac 2400
gacaagcgtg tatggagcga gtgagttcaa agcagaaaagg gaggagaagc aggcaagggg 2460
ggaggctgtg ggtgacacca gccaggaccc ctgaaaggga gtggttgttt tcctgcctca 2520
gccccacgct cctgccgggc ctgcacctgc tgaaccgtc gatgttggtg ccagggtgcc 2580
acctgggaag gatgctgtgc agggggcttg ccaaactttg gtgggtttca gaagccccag 2640
gcacttgtgg caggcacaat tacagccctt ccccaaagat gccacgtcc ttctcctgga 2700
acctgtgaat gtgtcaccgc caaggcagag gctggtgaag gctgcagggt gaatcacggc 2760
tgccagtcag ccgatcttaa ggtcatcctg gattatctgg tgggcctgat atggccacaa 2820
10 gggtccttag aagtgagaga gggaggcagg ggagagtcag agaggggacg tgagaaggac 2880
cactggccac tgctggcttt gagatggagg aggggggtccc cagccaagga atgggggcag 2940
ccgctccatg ctggaaaagc aagcaatcct ccccggtcct gagggcacac ggccctgccc 3000
acgcctcgat ttcaggccag tgggacctgt ttcagcttcc cggcctccag agctgtaaga 3060
tgatgcgttt gtgttcagcc actaagctgc agtgattcgt cacagcagca aatggaatag 3120
15 cagtacaggg aaatgaatac agggacagtt ctacagtgga ctctcagccc acccctggg 3179